2007

The relative influence of information assurance mechanisms on content integrity in peer-to-peer networks: a field study

Steven Michael Collins
Iowa State University

The relative influence of information assurance mechanisms on content integrity in peer-to-peer networks: a field study

by

Steven Michael Collins

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Information Assurance

Program of Study Committee:
Amrit Tiwana (Major Professor)
Sree Nilakanta
Steven Russell

Iowa State University

Ames, Iowa

2007

Table of Contents

Chapter 1. Introduction

1.1 Introduction

June 1st, 1999 was the beginning. A visionary named Shawn Fanning, along with two of his cohorts, released a program that would help kick off a peer-to-peer networking revolution. The program was Napster, and it would soon become the staple of the file sharing experience all over the world. Napster possessed a user friendly graphical user interface (GUI), a centralized server of lists of connected systems, and a specialization in audio files. End users enjoyed the program for its ease of use, but the introduction of information assurance mechanisms was largely not pleasant.

The absence of information assurance mechanisms in peer-to-peer networks did not become a pressing issue until later. The larger the networks grew the more content integrity began to become a problem. Users were posting malicious files, creators were flooding the networks with false and fake files, and inaccurate information was becoming the norm. This has become important from an information assurance standpoint since peer-to-peer networking is becoming more mainstream and being used by large organizations. Having such problems will damage business reputations, hurt sales, and all together take peer-to-peer networking down a notch.

To combat this problem, the peer-to-peer communities began using different mechanisms to help ensure the integrity of their content. These mechanisms flooded onto networks in droves, some successful and some not. The unfortunate aspect is that the networks implemented them many at a time, so ascertaining which influence end users

perceptions were not gathered. It could be that user ratings are a make or break for some users, while individual user accounts work for others.

Peer-to-peer networks are not limited to Napster and Kazaa style networks. They have branched out to encompass information sharing, video and picture sharing, and software sharing networks. These new networks branched away from the pure peer-to-peer architecture to more a centralized architecture. This allowed a central location to be used in order to help maintain the content, at the cost of significant bandwidth usage.

Until this point research has not branched out in an attempt to view the information assurance mechanisms in peer-to-peer networks. As such, no one knows how those mechanisms will affect the downloading habits of the end user when they use specific peer-to-peer networks that employ specific mechanisms. This paper will look at the mechanisms employed in peer-to-peer networks and, using a conjoint methodological approach, determine the affect on the end users downloading habits of specific information assurance mechanisms.

1.2 Significance of the Research Problem

Peer-to-peer networks were first introduced to the mainstream public many years ago with a little network known as Napster. The original iteration of Napster was that of a centralized peer-to-peer audio network that focused on the swapping of mp3 files. Since Napster's inception, the peer-to-peer market has exploded. There have been countless networks introduced and countless more networks shutdown. Table 1-1 shows 100 different peer-to-peer networks in 4 different categories arranged into 4 different subgroups. This

figure is explained in more depth in section 3.3.1. The sheer growth of the peer-to-peer community has been astounding.

Unfortunately, along with the massive data growth of the networks have come many different problems. The corruption of content that is to be distributed on these networks is among the top of the problems. The corruption occurs many different ways. Content owners have been known to spread around their own “works” in a corrupted format. One of the most famous examples is Madonna. She took an audio file of herself degrading the downloader, named it after one of her popular songs, and propagated it around peer-to-peer networks. End users were so outraged by this action that they retaliated in the form of hacking her website.

Owners also distribute other different corrupted files on these networks. Often times an owner will put up a file of random noise, unintelligible bits, or a file that is unable to be opened, under the disguise of one of their popular works. However, content owners are not the only ones to be blamed for file corruption. End users also propagate works that are corrupt in nature. These files are corrupted much in the same way as content owners corrupt their files. Many times on the more nefarious networks, such as Kazaa or Morpheus, files are infected with Trojan Horses, Virii, and other forms of Malware.

Peer-to-peer networks are also plagued with copyright infringement problems, including lawsuits. Many of the most popular networks, including Grokster, Kazaa and Napster were shutdown due to federal and international lawsuits. The two major players in the lawsuit campaign were the Recording Industry Association of America and the Motion Picture Association of America. Corporations and end users have a very different idea about what “fair use” actually means under the Digital Millennium Copyright Act. The end user

would like to download media they pay for and put it in any format they wish. Corporations would prefer that this not happen, as this leads to the potential for monetary loss.

Another big problem with peer-to-peer networks is content trustworthiness. When you log into a peer-to-peer network, how do you know that both the network and the content of the network are trustworthy? Most likely the average user has seen quality control mechanisms and used them to determine where and when to download. Unfortunately, standardization of these mechanisms is not in place. This means that some networks offer many mechanisms while some offer few to none. The content may be trustworthy, but without these mechanisms the users would be clueless as to its trustworthiness. On the opposite side, just because a network employs many mechanisms does not make the content trustworthy. The problem then becomes differentiating between these mechanisms to determine which are more likely to make your content more trustworthy and which are going to make end users more likely to use your network.

While it is true that peer-to-peer networks are often plagued with problems, information assurance mechanisms are often employed to combat the problems. However, since currently there are no studies done to show what mechanisms actually help your network and help to put your downloader's minds at ease, how is a peer-to-peer network supposed to know which mechanisms to employ? This study aims to answer that question and show which information assurance mechanisms will help users trust the content and people of a network.

1.3 Significance of Information Assurance Perceptions in Peer-to-Peer Networking

Many different groups of people would be interested in quality control mechanisms of peer-to-peer networks. The first major group can be classified as the typical “peer-to-peer pirate.” This is the person who hops onto a network and illegally downloads files. Users of Napster, Kazaa, Grokster, AudioGalaxy, etc. would mainly fall under this category. The second group would be content distributors. This group contains people who legally distribute their content to others via a peer-to-peer network medium. The third major group includes those “deal seekers.” These are the people who scour the peer-to-peer networks searching for the best deals on products or search for the best place to purchase a product. The final group would be the potential distributors. This is a fairly large group and includes the RIAA, MPAA, major television networks, bands, etc. The potential distributors are watching to see what sort of encryption methods can be used to protect their content as well as the viability of using a peer-to-peer network to distribute their content in order to minimize cost and maximize exposure.

There are connections between the categories or peer-to-peer networks and peer-to-peer network interested parties. The interested parties are the users of the different peer-to-peer networks, which is what makes them have a vested interest in this research. The peer-to-peer pirate is the primary user of the audio and video peer-to-peer networks. They really do not exist on the software or deal networks since there really is no copyrighted material to download. The deal seekers primarily stick to the deal networks since the other 3 networks have absolutely nothing to do with deals. The potential distributors would be interested in the audio, video and software networks. The audio and video networks pose a copyright infringement problem, which is what draws these potential distributors to these networks. They are looking for ways to distribute their content in a fair manner via these channels.

Finally are the content distributors. The differences between content distributors and potential distributors are that the content distributors are actually utilizing the peer-to-peer networks right now. The content distributors primarily stick to the audio and video networks as a way to disseminate their content.

The implications of using a peer-to-peer network are far reaching and affect many different aspects of the computer world. Certain information assurance mechanisms may help make peer-to-peer adoption more widespread. This has the potential to help the distribution of content removing some of the burden from the host or provider. For example, look at a centralized peer-to-peer network such as YouTube. YouTube spends more than $1 million per month for bandwidth alone.[1] However, if YouTube were to switch to a pure peer-to-peer network model, this would alleviate much of the cost of serving the movies. Instead of YouTube serving every movie to every viewer, the viewers would be able to transfer the movies to other users thus becoming the responsibility of the end users to maintain the network. There are many other implications with this example that are not being taken into consideration, such as copyright infringement and actual hosting of movies on users machines.

The widespread adoption of peer-to-peer networks also has the potential for generating user driven content. This is specifically viable in Web 2.0. Web 2.0 is the concept of a second generation internet revolving around wikis, social networking sites, and communication tools. Sites such as Digg and Del.icio.us have been heralded as the beginnings of Web 2.0. They revolved around the idea of peer-to-peer networking, mostly

[1] http://www.forbes.com/intelligentinfrastructure/2006/04/27/video-youtube-myspace_cx_df_0428video.html

focusing on the centralized peer-to-peer design. The proprietors of these websites, their user bases, and anyone wishing to imitate their design ideas have a stake in the future of peer-to-peer to networking.

1.4 4 Types of Peer-to-Peer Networks

Peer-to-peer networks can currently be broken down into 4 different categories: deal, audio, video, and software. Deal networks are networks that focus on the dissemination of knowledge or the sharing of deals. Users congregate here to share the newest deals or to share news stories. Digg, Wikipedia and EBay all fall under this category. Audio networks are networks that allow users to share music, movies, TV shows, pornography, video games, and pictures. Users will congregate to swap files, usually illegally to many other people. Kazaa, Napster, and Morpheus are all audio networks. Video networks are networks that focus on the distribution of audio, video and picture media. The different between video and audio networks is that video networks are centrally distributed online, whereas audio networks are typically decentralized and require specific clients to access the network. Examples of video networks include Google Video, YouTube, and Del.ic.ious. Finally, software networks have the more narrow focus of open source software and distribution. These networks include Project Gutenburg and Skype.

1.5 Summary of Research Contributions

Peer-to-peer networking is an emerging technology in the telecommunications field. Peer-to-peer can help alleviate the burdens put upon centralized servers for file transfers, as well as offer avenues for content contributors that may lack funding for content hosting.

However, peer-to-peer networking is not without its fair share of problems, including content trustworthiness, content corruption, and a plethora of legal issues. To help combat these problems, some networks have implemented information assurance quality control mechanisms to help users determine if they would like to download a particular file or not. Past research has not delved into the realm of quality control mechanisms to study which, if any, quality control mechanisms actually influence the end users downloading habits on peer-to-peer networks. As such, through the use of the conjoint method, this study aims to provide information on which specific quality control mechanisms influence end user downloading habits.

Chapter 2. Literature Review

2.1 Peer-to-Peer Classifications

2.1.1 Network Topology Classifications

For knowledge peer-to-peer networks, two classifications are used: pure and hybrid (Tiwana 2003). A pure network is one that is server-less and involves the nodes of a network interconnecting to one another. The hybrid network is one that involves a centralized server through which the flow is initiated but can also allow for nodes to connect to one another. These two classifications encompass a large portion of peer-to-peer networking, but they fail to address pure networks that use supernodes (Androutsellis-Theotokis, Spinellis 2004). The problem in labeling these networks then becomes deciding upon a third definition in which a pure peer-to-peer network that uses super nodes would fit. Since the idea of an end user being the super node would negate the centralized server while at the same time negating a pure peer-to-peer model, it leaves us trying to create a third label. In a study of 100 networks conducted by the researcher (Table 1-1), all 100 networks can be classified into pure, hybrid, and supernode structures. This does not, however, mean that these are the only three classifications that exist for peer-to-peer networking but merely the most commonly used that a peer will encounter.

2.1.2 Content Classifications

Peer-to-peer networks are able to deliver a wide variety of content. Blizzard Entertainment uses peer-to-peer technology to deliver game updates and patches for their widely popular World of Warcraft videogame. YouTube helps deliver 70 million videos to its

user base everyday.[2] Little Timmy is even using peer-to-peer technology when he logs into Morpheus of Kazaa to download the newest Christina Aguilera single.

Peer-to-peer networks come in a wide variety of distributions serving up a variety of content types. So how can we categorize these peer-to-peer networks based on the content that they distribute? Based on my research there has been little to no attempt to make these classifications. As such, I have no comparison base.

A survey was done of 100 networks, conducted by the researcher (Table 1-1). These networks were put into four general categories: video, audio, software and deals. It was thought that these four categories would encompass the peer-to-peer networks that were studied. From here 100 networks were identified, their content distribution types were identified, as well as four other categories that are not being considered for this specific section. Within the four categories, the content distributions were beginning to broaden and break free of the shackles of the previous labels.

Each category was then looked at to determine what types of content were actually being distributed. Based on these types of content, changes to all four names were proposed: deal changes to knowledge, audio changes to pirate, video changes to media, and software changes to open source.

Deal networks did focus on peers sharing their deals. However, the category goes much broader than that. Deal networks also focused largely on human to human interaction and knowledge transfers. Sites like Wikipedia and Digg focus on users sharing knowledge and news with one another, whereas sites like EBay and Slickdeals focused more on users

[2] http://www.youtube.com/t/about

sharing deals with one another. Based on this, the term “knowledge” has been substituted for deals to give a better encompassing of the overall network.

Audio networks, such as Kazaa, Napster, and Morpheus, are so much more than simple devices to share audio files. They encompass a large variety of files including: games, movies, music, text documents, serial numbers, pictures, and TV shows. While it is true that these networks do share legal files, the vast majority of what users share on these networks is illegal. Due to the illegal nature of these files and the fact that they do not focus strictly on audio files, the term “pirate” network is being used. This term fits these types of networks as the legality of the networks and the files that they housed have been challenged time and time again, leading to many networks being taken down permanently through court injunctions.

The video networks also focus on a lot more than just video files. They specialize in video, audio, pictures, bookmarks, and TV. One difference that should be noted between the video and the audio network is that the video networks have a hybrid peer-to-peer type that relies on a centralized server through which users upload and download content. As such, the term video cannot encompass this category. Instead, these networks should be referred to as “media” networks, since the term media will encompass what they offer.

Finally, there are the software networks. These networks seem to strictly focus on the open source community and legal file distribution. Since the open source software is the focus of these networks, it makes sense to rename this to “open-source” network.

These new network definitions do a significantly better job of capturing the essence of the types of peer-to-peer networks that were observed. Also, these new names better describe what sort of content is being distributed through the network without leaving specific content out. It should be noted that these names are not the be-all-end-all of the peer-

to-peer networking world. Instead, they encompass the networks that were observed as well as other networks of similar content. This is to say that there are not networks that exist outside the bounds of this naming scheme, but simply that those networks were not observed.

2.2 Trust and Reputation in a Peer-to-peer Network

One of the main problems that peer-to-peer networks suffer from is trust. How can you trust the file you are downloading, and how can you trust the person that is offering the file? For all the downloader knows, the file is infected with viruses, malware, spyware, Trojan Horses or worse. Peer-to-peer networks often protect the anonymity of its users. To help protect this anonymity, the networks use opaque identifiers for shared information. However, the downloader and the seeder's IP addresses become apparent once downloading beings (Damiani, Vimercati, Paraboschi, Samarati, 2003).

As it stands right now, there are currently three ways for trust to be given. First, a peer could give itself a rating based on what the peer feels he or she is contributing. For example, I could give myself a rating of 8 out of 10 if I felt I was trustworthy. This is probably not the best system as it can and will be easily exploited. The second way for trust to be given is by a trusted third party. For example, a certificate authority is a trusted third party responsible for giving out certificates. The Better Business Bureau is also a trusted third party. The final way trust can be given is from other peers. Peers can rate each other based on quality and relevance of the content they are providing. (Wang, Vassileva)

LOCKSS is a system proposed by Parno et all in 2004. In a LOCKSS system, the computers each house copies of the information in an ad hoc peer-to-peer network structure. Multiple copies of these files make the corruption of data significantly harder, if not near

impossible (depending on network size). Reputation management is something, however, that LOCKSS chooses not to discuss or implement. A node can easily abuse a reputation management system. The example given by Parno et al, is that a user can build up a positive reputation over a period of time and then leverage that positive trust into something malicious. The node could flood the network with corrupted files, pass off Virii, and many other things through leveraging their positive reputation on a peer-to-peer network.

2.2.1 Reputational Trust

The researchers at Stanford University developed a reputation based management system that is used for peer-to-peer networks called EigenTrust. The basic idea of EigenTrust is to assign each individual node within a network a global reputation rating (or value). These reputation ratings are then used to determine whether or not peers will download from another peer.

The basis of EigenTrust is to address 5 issues within a peer-to-peer reputation system: self policing, maintain anonymity, not assigning profit to newcomers, minimal overhead, robust to malicious collectives (Kamvar, Schlosser, Garcia-Molina, 2003). The first issue is that the system be maintained/enforced by the users within the peer-to-peer network and not by a neutral third party. Secondly, the reputation needs to be tied to one of the opaque identifiers to maintain the anonymity of the user. This means that reputation would be tied to something like a peer's user name rather than the peer's Machine Address Code (MAC) address. Third, the reputation must be earned and not given. That is to say that simply by a user changing their use name they should not be given a clean slate with which they can immediately take advantage. Instead, their reputation must be earned through how they act

over time. The forth issue is that of overhead. This means that the reputation system needs to put as little strain as possible on the actual infrastructure of the network and not take large amounts of storage to house. Lastly, the reputation system needs to not allow a united group of peers to game the system. If a group of people were to work together and give themselves all high ratings, they could theoretically give themselves high global rankings without actually earning them, and this cannot be allowed.

EigenTrust is an algorithm that was developed to specifically work through these five issues. EigenTrust is a very complex and intricate algorithm that takes many different factors into account. A simple explanation is that the algorithm takes in information from other peers and uses that to calculate a trust based rating. That rating is then normalized so that united groups of people cannot collaborate to defeat the system.

Why not simply implement the algorithm in all peer-to-peer networks? Implementation of this algorithm is a strong idea with a few small exceptions. This would, by no means, be the end of trust issues within your network. Users would still have the ability to spread malicious files with a high trust rating. A user could be a member of the network and gain a high trust rating through decent acts and then turn around and exploit that high trust to spread malicious files. However, that high trust would be quickly diminished but at the potential cost of compromised computers. Even though there is a feature built into the algorithm to prevent groups of friends from artificially inflating their ratings, it would still be possible. If large enough groups collaborated they could, in theory, inflate their own rankings.

2.2.2 Fuzzy Reputation Aggregation

The fuzzy logic reputation aggregation system was developed at USC specifically for use in e-commerce applications. The goal with the fuzzy system was to create a non overhead intensive accurate reputation management system that can be deployed in e-commerce peer-to-peer networks and potentially non e-commerce peer-to-peer networks. The system also tries to weed out untrustworthy opinions from users who are simply trying to give black marks to peers with good reputations (Song et al).

The fuzzy logic reputation aggregation system had three design characteristics: considers the unbalanced transactions among users, the more a person is involved with the system the more their reputation should update (as in super users update more often then small users), and evaluating the large transactions more often than the small transactions (Song et al).

The FuzzyTrust system aggregates the local trust scores of all peers and then gives out a global reputation for the peer. The aggregation function considers three factors when determining reputation scores: peer's reputation, the transaction date, and the transaction amount (Song et al). Fuzzy inference rules are also implemented. The rules can be very small (less than 5) to more than a few hundred depending on the size of the network and the amount of aggregation that is needed.

The overhead on the FuzzyTrust system is significantly lower than that of the EigenTrust system. FuzzyTrust averages 17 messages sent per peer, whereas EigenTrust averages 73 messages sent per peer. This helps to alleviate the inherent problem of super users contributing to most of the transactions with small users contributing significantly less. FuzzyTrust is also very scalable for large peer-to-peer networks, requiring 180,000 messages

for 10000 peers. EigenTrust would require 580,000 messages for those same 10000 peers. This makes FuzzyLogic much less on overhead and less problematic.

2.2.3 P2PRep

The concept of P2PRep is to allow a peer to query other peers in order to get reputation information about the peer that is offering content to download (Damiani et al). A peer would see a file to download offered by multiple users. The peer would then send out a poll request to other peers asking for their opinion(s) on the users with the content for download. Once the peer has their poll opinions back, the peer can then choose which user to download from based on the responses from other peers on the network.

P2PRep also attempts to overcome security vulnerabilities. The authenticity of those who are voting and those who are acting as content offerers, the quality of the polling, and the ability to detect “dummy” systems that are trying to cheat the system need to be identified. P2PRep also keeps track of the peer’s history through a servent_id, which helps maintain authenticity. The servent_id does not compromise anonymity. In order to determine if the votes by peers are unique and authentic, P2PRep will look at messages to determine if they are likely to have come from the same user or if they would in fact be unique. The ones that are determined to have come from the same user or group of users are discarded.

With P2PRep there are two ways of doing the polling. The first way is for the respondents to not provide their servent_id. The second way is for the respondents to provide their servent_id. If they provide the id, P2PRep can then take into account the peers history when calculating reputation. So the better approach is to have the peers provide their

servent_id to help maintain authenticity as well as to help calculate more accurate reputations based on a lot more information.

2.2.4 Peer-to-peer Incentives

An adaptation of prisoner's dilemma approach to peer-to-peer networks was proposed in 2005. Some challenges that are unique in peer-to-peer networks are: relatively high turnover rate coupled with large populations, interest is asymmetric and allowing users to be constantly changing their identities (Feldmen et al, 2005). The system that is proposed consists of four properties. The first property states that optimal utility is the cause of overall universal cooperation, while rewarding those who exploit the cooperation of others. The second property states that if one peer wants a service from a second peer but does not have anything the second peer wants in exchange, the first peer should still be able to take advantage of the service the first peer has to offer. The third property says that if someone defects against another peer, the peer who was defected against should not be able to find out the defectors identity. The fourth property says that peers should have the ability to enter or leave the system independently as well as be able to alter their behavior at random.

Prisoner's Dilemma	Man 2 Does Not Talk	Man 2 Talks
Man 1 Does Not Talk	Lesser sentence for both	Man 2 goes free, man 1 gets a more harsh sentence
Man 1 Talks	Man 1 goes free, man 2 gets a more harsh sentence	Both get a moderate sentence

Table 1-1 – A depiction of the prisoner's dilemma. Shows 2 different people with 2 different choices, and the resulting outcome when the decisions are made.

This sets up a matrix, where two peers have the option to cooperate or defect. Based on the decisions of the two peers, different payoffs are received. If both cooperate, the payoff will be higher than if they both defect. This system has obvious problems. When you give someone the opportunity to defect, or act negatively, without repercussions that extend into the real world then there is really no reason to choose to cooperate. If a peer can still get the service that they desire anonymously and without providing something in return to another peer, that peer will typically choose to do so. Traitors are also a problem. A traitor would be classified as someone that builds up a high reputation rating only to turn around and defect before exiting from the system. This means that someone could build up a high rating through good work only to turn around and flood a network with malicious files or act maliciously on a network with their high rating. Since all of this is done anonymously, there really are no repercussions to this either.

2.2.5 Reputation System Problems

Resnick et al, in 2000, proposed three significant problems that currently exist within the current reputation systems of peer-to-peer networks. The three problems are: eliciting, distributing, and aggregating. Eliciting suffers from three problems: apathy of users, negative feedback elicitation, and honest reporting. Distributing suffers from name change problems

and the lack of portability. Aggregating has the problem of the actual aggregation of the information and the displaying of the information.

Elicitation problem one relates to the apathy of users toward actually using the reputation system. Most feedback systems require you to fill out a form in order to leave feedback on a transaction that has occurred. These forms vary from very easy and short to very difficult and long. However, no matter what the form is, there is little to no incentive towards taking the time to fill out the form. EBay is a great example of this. On EBay once a transaction has been completed there is a short and simple form that you are required to fill out if you want to leave feedback. The incentive here is that if the seller leaves good feedback for a good transaction, the buyer will then take the time to leave good feedback as well, or tit-for-tat.

Elicitation problem two deals with the idea of eliciting negative feedback. As reputation and feedback systems have evolved, unfortunately they have evolved in such a way that retaliation is a large problem. On EBay if a seller leaves negative feedback for a buyer, the buyer often times retaliates and leaves negative feedback for the seller even if the seller has done nothing wrong. This makes most people think twice about leaving negative feedback and often times results in the feedback not being left. It is also a fairly common practice for people to work through the problem before leaving the negative feedback. This can bias the reputation system since no one would know that there was ever a problem.

The final elicitation problem is that of honest reporting of feedback. Many problems arise here including buying positive feedback, blackmail, and collaboration of users to inflate their rankings. Users could conduct false transactions with other users with the intention of leaving positive feedback that they have been paid for. One user could threaten to post

negative feedback for another user if positive feedback is not left. This tarnishes both people's reputations. These problems make the reputation system meaningless since the actual ratings of the users would be inaccurate.

Users' pseudonyms are garnered when a user registers for a site. If that user decides to change their user name, their past feedback will not go with them. This can have positive or negative side effects. If a legitimately good user with good feedback wants a name change, it is often difficult to have their old feedback transferred to their new pseudonym. However, if a malicious user with negative feedback wants to change their pseudonym they essentially get a clean slate and a fresh start.

Reputation systems are also proprietary or lacking in portability. This means that a user's EBay rating and Amazon.com ratings are different, and a user has no way of knowing if you do malicious things on one but not the other. As it stands now there is not a wide spread acceptance from the general public for sites that do offer a way to track all of the reputations a person may have.

Finally, there is the problem of feedback and reputation aggregation. Some sites choose to display your average rating while some display your overall feedback. Currently, these fail to answer important questions such as the value of the transactions, why negative or positive feedback was left, and the reputation of the person that left the feedback. All of these are important things to consider when looking at someone's overall reputation, but currently none are taken into account.

Hypothesis 1 – The total number of previous downloaders that have rated a specific contributor will have a positive influence on peer-to-peer network content information assurance.

Hypothesis 2 – The reputation of a contributor, as given to him by the peers in a specific peer-to-peer network will have a positive influence on peer-to-peer network content information assurance.

Hypothesis 3 – How extensively a contributor reputation and/or rating system is used in a specific network will not have a positive influence on peer-to-peer network content information assurance.

2.3 – File Quality Characteristics

2.3.1 Certifying Data

When talking about the file characteristics on peer-to-peer networks, the idea of integrity of the file comes up. Some networks allow the nodes to be able to verify the integrity of a file. (Castro et al. 2002) First, a peer wants to upload a file to a network. Once the upload process has been completed the peer will compute a hash for the file and make the hash known. A different peer then connects to the first peer, downloads the file that the first peer is sharing, and computes the hash to verify the integrity of the data.

Another example of certifying data is an anonymous cryptographic relay. An anonymous cryptographic relay is made up of four parts: the client, the publisher, the forwarder, and the storer. The concept behind the relay is for a publisher to decide that they want to upload a file. The publisher will chose different forwards to send parts of the data. The forwarders then decide who the storer(s) will be and send all the parts of the file to that individual. The storer(s) reassembles the data into the completed file. Once this has taken place, the publisher will then delete their copy of the content and let the rest of the network know what the file name is, and informs the network of who the forwarders are. (Androutsellis-Theotokis, Spinellis, 2004)

To download a file, the client would need to get in contact with the forwarders. The forwarders let the storer(s) know who is looking for the information. The storer(s) will decrypt the data and send the files back to the client.

There is the potential for attacks in a scenario like this. The forwarders could act as a man in the middle and send the storer different data that has been corrupted. The storer could also alter the end product and corrupt the file. Since the original author does not compute a cryptographic hash for this file, there is potentially nothing to stop this sort of behavior.

Freenet is a free software application that touts an almost completely anonymous peer-to-peer networking experience. They are able to maintain an almost completely anonymous experience through a few methods. First, the information is sent encrypted. Secondly, all traffic is routed through different nodes so that knowing who has requested a file, or even what that file is, is extremely difficult. Users give up a user selected portion of their bandwidth and their hard drive to support the network. However, Freenet controls what content is actually stored on the user's computer in an encrypted manner so that the user has no knowledge of the data. (http://freenetproject.org/whatis.html)

This could obviously lead to problems. Content integrity could be compromised through the spreading of malicious files, and there would be no repercussions. Since there is virtually no way to know the content providers, people would, in theory, be free to spread around the malicious file of their choice.

In order to prevent the corruption of files on a large scale or the removing of content entirely, Parno et al proposed a system known as LOCKSS. LOCKSS stands for "lots of copies keep stuff safe." The idea behind LOCKSS is to propagate your files around the peer-to-peer network so that multiple copies exist making the deletion of an entire piece of content

significantly more difficult. The goal is to connect libraries with one another in an ad hoc peer-to-peer network fashion to allow for long term archival storage. Each LOCKSS computer will keep a copy of the information in their memory, and they will also participate in “opinion polls” to help detect and repair the attempted corruption of data (Parno et al, 2004). Since the information is then housed on multiple computers and the computers use opinion polls to help prevent data corruption, an individual would need to subvert a large portion of the computer in order to actually do any damage. This system relies on average computer hardware and not the use of super computers.

2.3.2 Trusted Reference Monitor

In “Enhancing Data Authenticity and Integrity in P2P Systems” Zhang et al propose a trusted reference monitor, or TRM. A TRM can help to monitor and verify the authenticity of the data that is being passed back and forth between nodes on a peer-to-peer network. TRM would use credentials of the hardware in order to digitally sign information that is being sent between nodes. The TRM would run in the operating system on the user’s space to maintain access control policies.

The TRM works by querying the providers’ TRM to make sure it is in a valid and non corrupted state. The providers TRM will send messages to a secure kernel requesting information, which is digitally signed and passed back to the TRM. The TRM also digitally signs data that is combined with the secure kernels information, and then passed back to the requesting TRM. If the values of providers’ signatures are valid, the requesting TRM will then pass along access control policies and some configuration policies. The two computers (or more) will then have a connection established.

TRM is also built with low overhead and minimalist machines in mind. When running 2000 queries without security measures, the queries took approximately three seconds, whereas with the TRM in place the queries took about 6.4 seconds. So with TRM the query time would double. However, it should also be noted that the machine TRM was tested on was a Pentium 3 600 MHz machine with 256 megabytes of RAM. A higher end computer, or a computer with better specifications, would have improved query times. Overall, however, the TRM query takes approximately double the time of a security-less query, when looking at 500, 1000, 2000, and 4000 queries.

2.3.3 Free-Riders and Incentives

A study on the Gnutella peer-to-peer network showed that 70% of nodes on the network provided zero files, and that the top 1% of the nodes accounted for 37% of all files shared on Gnutella (Feldman, Chuang, 2005). However, when the same study was conducted in the year 2005, it showed that the number of people who contributed zero files had risen from 70% to 85%, an increase of 15%.

“Free-Riding” is a common phenomenon in peer-to-peer networking. Free-riding is the concept of a node on a peer-to-peer network maximizing his own utility at the expense of others, meaning that the node will take from everyone else without actually providing anything to other nodes in return.

Feldmen and Chuang propose three different incentives to encourage cooperation in a peer-to-peer network: inherent generosity, monetary payment schemes, and reciprocity based schemes. Inherent generosity is the idea that a node will share their files without other nodes based specifically on the idea that the node will gain utility through the act of file sharing

itself. This means that a user will make a determination if they will share files based on what the cost is to share with the system. Monetary payment schemes focus on the idea that the recipient of the services being offered will pay for those services. The problem with this model is trying to keep track of all of the transactions and micro-payments in the system while still trying to keep some semblance of anonymity present. The final incentive is that of reciprocity. The idea here is to keep a history of the behaviors of nodes in the network which other nodes can use to decide whether or not they wish to provide service. This breaks down into two different schemes: direct and indirect. In the direct model node X makes the determination on whether or not to serve node Y based only on how much service node Y has provided node X. However, in the indirect model node X takes into account how much node Y has provided to the rest of the network.

Incentive schemes provide a way to encourage people to contribute to the peer-to-peer network(s) they are a part. Consequently, the incentives can still easily be manipulated and worked around. By offering monetary incentives malicious users will find ways to gain the monetary payments without actually doing work or by lying about the services that they provided. With using a reciprocity based approach you have the problem of exclusion. User A may have a file that user B wants but will not share it with user B until user A gets something in return which user B may not be able to provide. This could be a hindrance to the network and instead of encouraging growth will actually hinder the growth.

Hypothesis 4 – How much confidence an individual has that a particular peer-to-peer network has not been flooded with fake/malicious/corrupt files will positively influence the peer-to-peer network assurance.

Hypothesis 5 – The number of previous downloaders who have rated a particular file will positively influence peer-to-peer network information assurance.

Hypothesis 6 – The quality of the ratings for a file (how high or low the previous downloaders have rated that particular file) will positively influence the peer-to-peer network information assurance.

2.4 Authentication and Identity Management

It is often the case that authentication and identity management mechanisms in peer-to-peer networks take a back seat to other categories of mechanisms. In a network where access control mechanisms have been implemented, those mechanisms will be guided by the use of discretionary access control (DAC) (Androutsellis-Theotokis, Spinellis, 2004). DAC is the idea that if a user has access to something, that user can then pass on their access to another user at their sole discretion. This is obviously not a secure method as your network has un-trusted clients.

A major security threat to peer-to-peer networks is the idea that a single user can use many pseudonyms as they can throw out malicious files under different names in an attempt to trick other peers. This will often be found in peer-to-peer networks where the populations are constantly coming and going (Androutsellis-Theotokis, Spinellis, 2004). This problem was discussed by Douceur (2002), where it was determined that a central certificate authority or a central identification authority would be required to minimize this attack.

Some networks have decided to require users to register themselves offline. One such group is Intergroup. They require a user to sign up off of the network to obtain an X.509 certificate. With this certificate the individual is able to authenticate themselves to the network and actually use the network. The service by which the X.509 certificates are signed and distributed is Akenti. Akenti is certificate authority that will issue the certificates to the users that will allow them to connect to the network. This idea can still be circumvented by

an individual signing up using false information which can still lead to the spread of malicious files around a network. However, by requiring registration and verification through the use of X.509 certificates a network can assume a reasonable amount more security than a network that opts not to use such certificates.

Another approach to access control comes in the form of OceanStore. Oceanstore has two different design goals: ability to construct from an un-trusted infrastructure and the ability to support roving data (Kubiatowicz et al, 2000). The access control schema of OceanStore has two different restriction types: reader and writer. The reader restriction helps to prevent unauthorized reads, so the data is encrypted. This encrypted data consists of non-public information and the key is distributed to those with only read permission. If someone revokes data, replicas must be deleted from the network, or the data must be re-encrypted with a new key which would be distributed again to those with access. The problem, however, is that a person that has read access could still read old cached copies or if a server decided not to re-key the data. The writer restriction helps to prevent non-authorized writes. Every single write must be signed, so that clients and servers can check the signature against one from the access control list (ACL). The owner of the data can choose the ACL. There seem to be no security problems with the writer restriction.

So the OceanStore idea helps to keep the integrity of the data through the reader restriction and the writer restriction, whereas the identity of the content writer can be verified through signatures that are compared with known ACL's.

Pastry uses unique 128 bit identifiers in order to identify the peers of a network, known as a NodeID (Lua et al 2005). The NodeID is used to show what the peer's location is within a node circle. When a peer joins the network, it is randomly assigned this NodeID and

is assigned so that the identifiers are evenly distributed. In order for messages to be propagated through the system, Pastry will send the message from a peer to the node whose NodeID has a value closest to the given key. IP addresses are also kept in routing tables allowing for users to be authenticated via their IP address.

2.4.1 Key Management and Certificates

In “Mobility Helps Peer-to-Peer Security” (Capkun et al, 2006), they identify two different scenarios. The first scenario involves a centralized authority. This authority is responsible for the management of membership into the peer-to-peer network. Each specific node on the network is given a signature that is signed by the authority that is used to bind its identity to its public key. The second scenario assumes that the mobile network is fully self organized, where no PKI is present, and there is no central authority. Their proposed system uses cryptography and key management in order to establish identity and authentication. In the first scenario, the users are not aware of the security relationship establishment since the central authority would take care of it. Whereas, for the second example, there is no central authority the user is going to need to be consciously aware and make decisions on the security.

The proposed mobility based approach takes far less overhead than a lot of other approaches. The most that would be needed for the approach is a central authority. There would be no preloading of key pairs, no online key distribution center would be needed, and no complex security protocols would be required for full functionality. So the mobility-based approach would be superior in that there is less in terms of setup, less in terms of overhead, and more ways to verify security of those whom have already met.

2.4.2 Herbivore/CliqueNet

Herbivore can be described as an anonymous, scalable, tamper resistant protocol that can be employed in peer-to-peer networks (Goel et al). Herbivore consists of three different parts: communication endpoints are completely hidden, number of users is not influential since it can scale to support them, and the bandwidth and latency are not problems since it deals with them efficiently. The protocol works on top of your existing un-trusted peer-to-peer network based on the idea of a dining cryptographer network to ensure that the destination and the origin of messages cannot be garnered.

In order to achieve efficient and anonymous message sending, Herbivore breaks down the nodes of a peer-to-peer network into a specific size anonymous clique. That is to say that based on the number of participants and the number of nodes needed to maintain the desired degree of security of the system, Herbivore will automatically divide up the nodes to keep this security. When new nodes join the network or when the cliques become too large, Herbivore automatically segments the nodes into new cliques of nodes. Due to the way Herbivore segments off the network and the way that it sends messages through the network, Herbivore is resilient to the following attacks: collusion and occupancy, Sybil, topology, intersection, statistical, coordinator, exit, and denial of service.

Through tests done with wide area networks with Herbivore, it has been shown to achieve low messaging latency and high anonymous bandwidth (Goel et al). The more nodes in a clique and the more messages, the higher the message latency becomes. However, even with a clique size of 40 and four messages being sent at once, the messaging latency will only slightly more than double, going from approximately .4 seconds to slightly over 1

second. When there are four messages being sent in a clique size of 40 nodes, the bandwidth is still high enough to be able to use a web browser and stream audio and video.

With Herbivore you get a significantly high level of anonymity which could cause a lot of problems. With that high level of anonymity, there is absolutely no recourse for what you do on the network. You can offer up all sorts of malicious files to other users without being able to be traced, or you could violate copyright laws without repercussions. Couple this with no authentication to use the Herbivore service and no registration, and you have the makings of a network that would have severe trust problems.

Hypothesis 7 – The extent to which a contributors' identity can be verified in a specific peer-to-peer network (e.g., through requiring registration, valid email addresses, or other forms of identification) will not positively influence the peer-to-peer network information assurance.

Hypothesis 8 – How extensively a file rating system is used in a specific peer-to-peer network (e.g., detailed comments left by previous downloaders) will not positively influence the peer-to-peer network information assurance.

Hypothesis 9 – The ease of use of the file and contributor rating system on the users making download decisions will not positively influence the peer-to-peer network information assurance.

2.5 The Negative Side of Peer-to-peer

Peer-to-peer networking has high potential for fast and efficient content distribution. The distribution does not come without a seedy underbelly though. Negative aspects of peer-to-peer networks should be apparent the moment they are looked at. Members spreading incorrect files, virus ridden files, malicious users, copyright infringement and an almost total lack of accountability are some of the major problems with peer-to-peer networking.

2.5.1 The Presence of Malicious Files

A fairly famous artist by the name of Madonna was tired of users pirating her music, so she took the law into her own hands. She created an account on a popular network and uploaded mp3 files that consisted of her screaming at and belittling fans under the guise of her actual songs. The hacker community at large despised her for what she did and set themselves upon her. Her website was severely defaced for a period of time, and a large portion of her incorrect files were removed. Large organizations such as the RIAA and the MPAA have been known to upload corrupted files under the guise of popular newly released content as well as uploading fake Torrent files to BitTorrent search engines.

These are just two examples of users spreading incorrect files. The examples do not take into account the average end user but organizations and people who are purposely trying to game the system. An actual end user is far worse. Incorrect files are propagated throughout peer-to-peer networks at an alarmingly fast rate causing them to be difficult to eradicate.

2.5.2 Copyright Infringement

It seems that every week a new article is being published in one of the leading news publications around the country about another individual being sued by a major organization (typically the record labels and movie studios). Copyright infringement plagues the vast majority of peer-to-peer networks causing a slew of lawsuits and Digital Millennium Copyright Act notices to be distributed from internet services providers to end users.

In 1999, the first of the peer-to-peer networking copyright infringement lawsuits was filed against Napster. The Recording Industry Association of America (RIAA) filed suit in Northern California claiming that Napster facilitated the growth of a black market for music

file trading. The lawsuit sought $100,000 per copyrighted song that was traded on its network. At the time Napster had approximately 200,000 songs which brought the lawsuit to a total of 20 billion dollars. An injunction was granted against Napster so that in July of 2001, they were forced to shut down their network. The penalties did not end there as Napster was also forced to pay over $30 million for past infringements.

The copyright infringement lawsuits did not stop there. Kazaa, Morpheus, and even The Pirate Bay have all seen attempts at having their networks shut down. However, copyright infringement notices seem to only extend to two of the four types of peer-to-peer networks. The pirate network and the media network have been trying to find ways to avoid infringing material for years while the open-source and knowledge communities have had little to no problems with infringement.

2.5.3 Malware and the Trojan Horse

It is not overly uncommon to download a file from a peer-to-peer network and find that it has been infected with some sort of virus or spyware/malware. The average PC user is only slightly aware of any security risks that peer-to-peer networking presents, and when coupled with how insecure the PC platform of computers has come to be, malicious individuals can easily take advantage through infected files (Parameswaran, Susarla, Whinston, 2001).

A popular method of infecting a user's computer is to infect a file and name it after a popular download. A user will download this file and attempt to run it causing the infection to spread to the user's computer. Typically these infections come in the form of back door Trojan Horses. The Trojans will then be used to connect to a remote server which is typically

based in IRC. Next, the individual that infected the file can connect to the same IRC channel as the infected computer and issue remote commands. That computer can now be controlled remotely through IRC allowing the perpetrator access (Eric Chien, 2003).

In 2006, a group from the Computer Science Department at Indiana State University conducted an experiment in an attempt to deduce how many files on OpenFT and Limewire were infected with malware. On Limewire, 27,717 files downloaded contained some form of malware. Those 27,717 files accounted for 35.5% of the total number of files they downloaded. On OpenFT, 599 files were downloaded that contained some form of malware. The 599 files accounted for 3.4% of the total files downloaded. It should be noted that the files that were downloaded were executable file extensions and Microsoft Office file extensions. What this means is that a typical user is going to have a fairly significant chance of downloading a malware infected executable on Limewire (Kalafut, Acharya, Gupta, 2006).

2.6 Definition of Key Terms

Peer-to-peer Network - Peer-to-peer systems are distributed systems consisting of interconnected nodes able to self organize into network topologies with the purpose of sharing resources such as content, CPU cycles, storage and bandwidth, capable of adapting to failures and accommodating transient populations of nodes while maintaining acceptable connectivity and performance, without requiring the intermediation or support of a global centralized server or authority (Androutsellis-Theotokis, Spinellis 2004).

Peer – the entities that are connected in a peer-to-peer network (Androutsellis-Theotokis, Spinellis 2004). For the purpose of this study, the term peer will also be synonymous with "end user" and "node."

Supernode – nodes that function as dynamically assigned localized mini-servers (Androutsellis-Theotokis, Spinellis 2004).

Information Assurance - Measures that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation. These measures include providing for restoration of information systems by incorporating protection, detection, and reaction capabilities (Committee on National Security Systems).

Quality Control Mechanism – In terms of peer-to-peer networks, a quality control mechanism is a means of ensuring content or contributor integrity. For example, user profiles and file ratings are both quality control mechanisms.

Semantic Differential Anchor - Measures people's reactions to stimulus words and concepts in terms of ratings on bipolar scales defined with contrasting adjectives at each end (David R. Heise).

Malware – A piece of software that has been inserted unknowingly or unlawfully into a computer system in order to cause damage.

Trojan Horse – A program destructive in nature, that acts as a legitimate application. This program can send back personal information stored on your computer, your keystrokes or allow remote access to your computer.

2.7 Summary

Hypothesis	Definition
1	The total number of previous downloaders that have rated a specific contributor will have a positive influence on peer-to-peer network content information assurance.
2	The reputation of a contributor, as given to him by the peers in a specific peer-to-peer network will have a positive influence on peer-to-peer network content information assurance.
3	How extensively a contributor reputation and/or rating system is used in a specific network will not have a positive influence on peer-to-peer network content information assurance.
4	How much confidence an individual has that a particular peer-to-peer network has not been flooded with fake/malicious/corrupt files will positively influence the peer-to-peer network assurance.
5	The number of previous downloaders who have rated a particular file will positively influence peer-to-peer network information assurance.
6	The quality of the ratings for a file (how high or low the previous downloaders have rated that particular file) will positively influence the peer-to-peer network information assurance.
7	The extent to which a contributors' identity can be verified in a specific peer-to-peer network (e.g., through requiring registration, valid email addresses, or other forms of identification) will not positively influence the peer-to-peer network information assurance.
8	How extensively a file rating system is used in a specific peer-to-peer network (e.g., detailed comments left by previous downloaders) will not positively influence the peer-to-peer network information assurance.
9	The ease of use of the file and contributor rating system on the users making download decisions will not positively influence the peer-to-peer network information assurance.

Table 1-2 - Depiction of the 9 hypothesis and what each of the hypothesis means.

Chapter 3. Methodology

This chapter describes the research methodology that was used to test the hypothesized relationships. A field experiment using a conjoint research design was given to Iowa State University students.

3.1 An Overview of the Conjoint Method

Conjoint analysis is a multi-attribute judgment analysis technique based on Information Integration Theory that involves posteriori decomposition of the respondent's decision process (Tiwana et. Al, 2006). The conjoint research method consists of three different pieces. The first piece is the attributes. The attribute is what helps the respondent make a decision about a dependent variable. For this study, there are nine different attributes for which the respondent will use to evaluate two different dependent variables. The second piece is called the conjoint profile. A conjoint profile consists of all nine of our attributes where the attributes are given a value. The value for this study is bipolar and can either be "high" or "low." This study will consist of 12 different conjoint profiles where no two profiles are the same. The final piece is the overall utility and the part-worth utility. The overall utility refers to the value that has been assigned to the dependent variable. The study consists of two different dependent variables where the overall utility is measured on a scale from 1 to 9, where 1 is very low and 9 is very high. The part-utility consists of the nine different attributes and their contributions to the overall utility as a whole.

The purpose of the conjoint analysis method is to have a respondent make multiple judgments about the dependent variables using the attributes as their basis for judgment.

3.2 Criteria for the Choice of the Conjoint Methodology

Traditional research methods such as surveys only allow for one attribute to be the testing point, so when multi attributes need to be considered a different approach must be taken. One of the main factors that was considered when choosing the conjoint method was the ability to do a multi attribute analysis. Peer-to-peer networks traditionally use more than one type of quality control mechanism often times employing greater than five. The whole basis of this study was to deduce which of these mechanisms actually influences an end users decision making process. Thus testing one attribute would not have yielded results that would be of use. This study employs nine different attributes that the user passes judgment on to come to a final verdict about the dependent variable. So a conjoint approach will allow me, in tandem, to test nine different attributes and their influence on the two dependent variables.

In a traditional survey, a respondent is asked to consider one of their past experiences in order to pass judgment on the question(s) that is presented to them. The researcher was asking users questions based on peer-to-peer network. If the researcher was to choose a traditional survey method, this may raise an ethical or legal question in the respondents mind. They may only have had unsavory experiences with peer-to-peer networks. The respondent may not want these experiences to be public, or they may not want to admit to anyone despite the anonymity of the survey and the complete lack of repercussions from it that they have done something illegal. This would then cause the results to be skewed and inaccurate which would negate a lot of the work that was done. In order to eliminate this bias, the conjoint method was chosen. Since a conjoint approach does not ask a respondent to consider their

past actions but instead asks them to evaluate a hypothetical situation. This can help eliminate biases that may arise.

These were the two determining factors when trying to choose which research method to use. The conjoint method offered a more robust way to test the data and also offered a multi-attribute approach.

3.3 Phase I: Qualitative Analysis of Peer-to-Peer Networks

The first step taken in this thesis research was to come up with a list of peer-to-peer networks. The networks were originally broken down into four different categories: video, audio, software, and deals. Once these four categories were established a list was started. The list first contained five different peer-to-peer networks for each category. Along with the names of each network, the following categories were recorded as well: the type of content that the peer-to-peer network housed, the major quality control mechanisms employed, an approximate age of the network, size of the network, and the major problems with the peer-to-peer network. Once five networks from each category were determined the search was ramped up to find 20 networks for each category. Finding the networks usually meant searching around the internet or asking friends and colleagues. Once 20 networks in each category were determined the count was up to 80. From here 20 more networks were ascertained to make the total number of networks 100. Some categories have more networks than others based on ease of finding said networks and really how many were out there.

Once 100 peer-to-peer networks consisting of four different networks were determined, the process of refinement began. Refinement consisted of adding five more columns to the excel sheet: knowledge network, file trading network, content sharing

network, more than one, and none. These labels represented the type of network that the specific peer-to-peer network represented. For example, Kazaa is primarily a file trading network, whereas Google Video would be a content sharing network. So these 100 networks were analyzed to determine how they were used. When a network was associated with one of the five categories, a 1 was put into the box corresponding to that category.

Once this was completed, further refinement took place only this time the definitions of each of these categories was looked at. A definition of a peer-to-peer network from a paper was looked at and then compared with each of the four categories (Androutsellis-Theotokis, Spinellis, 2004). Each of the categories already had a definition given to it by myself. So each category was analyzed to determine how my definition and the paper's definition differed. From there a hybrid definition was developed to suit each categories needs. Also included in this table were comments based on each definition to help justify the new definition. (see appendix figure 1-2)

The list of 100 networks and their properties helped to determine information that went into the conjoint profiles and the dependent variables that the conjoin profiles would be analyzing.

3.4 Design of the survey

The development of the instrument to be given to respondents was broken into different phases. The first phase involved designing the survey. Different attributes needed to be identified that could be given the characteristics of "high" or "low." From the list of 100 peer-to-peer networks surveyed, common quality control mechanisms were observed. There were quite a few mechanisms in common between the networks and some mechanisms that

were only employed on a very small portion of the networks. The mechanisms that many of the networks had in common were compiled into a list. From that list, it was determined which could be associated with a ratings system consisting of “high” or “low.” Nine factors were then chosen: contributor reputation, # of ratings for contributor, contributor identifiability, contributor reputation system usage, file quality rating, # of ratings for this file, file rating system usage, P2P network trustworthiness, and ease of use of ratings.

3.4.1 Development of Conjoint Profiles

This survey contains 12 different conjoint profiles. Each of the 12 profiles depicts a non-repeated peer-to-peer network. Each of these networks has nine different attributes that contain various combinations of options that define the network. The attribute of contributor reputation referred to the rating that the contributor had as given to the contributor by other users of the peer-to-peer network. The attribute of number of ratings for contributor referred to how many of the previous downloaders had given a rating to the contributor. The attribute for file quality rating referred to how high the previous downloaders had rated that particular file. The attribute for number of ratings for this file referred to how many of the previous downloaders had rated that particular file. The attribute for P2P network trustworthiness referred to how much confidence you have that this P2P network is not flooded with fake and/or malicious files. The attribute of contributor identifiability referred to the extent to which the contributors’ identity could be verified in that P2P network (e.g., by requiring registration, valid email addresses etc). The attribute for file rating system usage referred to how extensively a file rating system is used in this particular P2P network (e.g., detailed comments left by previous downloaders). The attribute for contributor reputation system

usage referred to how extensively a contributor reputation and/or rating system is used in this particular P2P network. The final attribute of ease of use of ratings referred to how easy it is to use the file and contributor rating system to make download decisions.

In this study there were two dependent variables: For downloading this file to my computer, and what is the likelihood that you will download this file? Each was evaluated using a nine point semantic differential scale. The semantic differential anchor was a bi-polar scale. For the variable “for downloading this file to my computer,” the two anchors were “risks greatly exceed benefits” and “benefits greatly exceed risks.” The variable “what is the likelihood that you will download this file?” had the anchors of “very low” and “very high.”

Independent variables will have correlations; however that is beyond the scope of this research study. In the research design it is assumed that the independent variables are orthogonal (non correlated).

Since peer-to-peer networking often times involves the trade of illegal files, three assumptions needed to be made by the respondents. The first assumption was that the user of the peer-to-peer network is completely anonymous. This means that they cannot be traced and what they download is not able to be linked back to them. The second assumption is that the file they are downloading is completely legal, the file takes 10 minutes to download, and the file is 500 megabytes in size. By making the file legal, the respondent can take the ethics and legality problem out of their decision. With a file that is 500 megabytes in size and only taking 10 minutes to download, the respondents will spend equal time downloading the file. Using high speed broadband versus a dial up connection to downloading, this file was removed by adding this assumption. The third and final assumption is that the peer-to-peer

network has existed for five years. This means that the network has been operational for five years making it a fairly old peer-to-peer network.

Once the conjoint profiles were done, the dependent variables were determined, and everything was put together into a two page (1 page duplexed) instrument. Two measures were averaged in order to determine a score, and the variables were 2 ways of capturing the data..

The instrument underwent various rounds of testing. The original was proofread for grammatical and spelling errors. Format changes also occurred here to make the instrument easier to read. After various initial rounds of redesigns and changes, a rough draft was established. This rough draft was to be given a group of 10 individuals. It was thought that if a small group would pre-test the instrument, problems and confusions could be established and corrected before unleashing it onto the masses.

For the pre-test, ten individuals were chosen on a non random basis. The researcher chose to distribute this to individuals with either a technical background or who had experience using peer-to-peer networks. Of the ten respondents, only one individual had never used a peer-to-peer network. The respondents were given a half page instruction sheet printed on pink paper and one double sided instrument printed on white paper. They were then asked to read over the instruction sheet and mark places they felt were unclear, confusing, redundant, or unnecessary. Once they were done with this, they were asked to work their way through all 12 profiles. They were also asked to mark up the instrument in the same way that they marked up the instruction sheet. When the respondent finished, the researcher asked specific questions about what they circled. This was done in order to clear up confusion that may have resulted from poor handwriting or lack of information about

where the confusion came from. This information was noted by the researcher on their forms. Once he had issued to and received back the instruments from 10 different people, the results were recorded. The respondents' answers are as follows:

- #1 For "# of ratings for contributor" and "# of ratings for this file" – Are we supposed to assume that all of the ratings are good? For example, you can have a high # of ratings on E-bay, yet a decent amount of them can still be negative. The first time reading through the "risks greatly exceed benefits, benefits greatly exceed risks" it seems like both of them are on the positive side. Took a few read times reading through it to understand the exact meaning.

- #2 Print is too small for people who have poor eyesight. The small font makes everything squish together making it difficult to read. What are people who wouldn't ever use these networks supposed to do? Should they assume that they do actually download and put aside all prejudice? (note, user actually asked the researcher if the assumptions were accurate, or if he was just trying to trick them for some reason, so perhaps something about no deception?)

- #3 No problems understanding the survey and nothing was confusing despite a lack of technical expertise.

- #4 The reference card and the search results do not match up, as in the order of them is different. Where it says "based on this information and your past expertise," shouldn't that be past experience since the users are not experts? Why is there a contradiction in price? On the second line, it says a $30 video game or movie DVD, but we are supposed to assume that the file is legal. On the reference card, for "file rating system usage," the file should be all caps to be consistent with the search results

- #5 The reference card and the search results do not match up, as in the order of them is different. The legal download yet $30 cost seems to be a contradiction, especially when using Kazaa, Limewire, or BitTorrent. For number "# of ratings for contributor" and "# of ratings for this file", what is the numbers were high but the ratings were bad? Is the "# of ratings for…" supposed to be on a per user basis or on a per file basis?

- #6 The font is a little on the small side. The gender question on the back may not matter if only targeting MIS majors, since there are almost no females in the MIS program. Survey was straight forward and not confusing.

- #7 In the box "for downloading this file to my computer…" it seems confusing at first glance, since the second sentence seems ambiguous. Perhaps put #1 for the first question and #2 for the second question to avoid confusion?

- #8 A lot of these options are not available in peer-to-peer networks. This makes it hard to make a determination based from these. For example, Limewire does not offer a lot of the options listed. Make a decision based on hypothetical networks? Survey was straight forward and easy to understand.

- #9 On the fourth line, there is a period, but on the fifth line it does not start with a capital letter and there is another period. What are the general risks and benefits a user is supposed to think about when answering the questions? Are they supposed to make their own assumptions? Overall the survey was a little difficult to understand at first. It took a second read through to figure out what you were looking for.
    - Fixes: Maybe not have the tables so close together? It can look incredibly daunting when you first look at it.

- #10 The line about "identifiably" is a confusing word, even after reading through the description what was actually meant was confusing. The two choices of majors, MIS and Non MIS, might be a bit too narrow.

It should also be noted that respondents names were not recorded, making this process 100% anonymous.

The information garnered from these ten surveys was then used to modify the survey to come up with the final version of the instrument. The font was changed to be a bigger size, and the dependent variables were reworded for less confusion. On the instruction sheet, the reference card section was changed so that the attributes would line up correctly with the attributes on the instrument.

3.5 Survey Administration

Once the final version of the instrument (Figure 1-2) and the instruction sheet (Figure 1-3) were established, respondents needed to be identified. The class that was targeted for the survey was a large core class for the business college. This class was specifically targeted for two reasons. The first is that the class is required for all business majors and deals

specifically with technology making it ideal to ask people about peer-to-peer networking. Dealing specifically with this class was appropriate in order to get a cross section of students with different majors that are all likely to have at least some experience with peer-to-peer networks.

Before the survey could be officially deployed, there were still a few more steps that needed to be taken care of. In order to ensure the survey accountability, each was individually numbered. Since there were 550 potential respondents, the surveys were numbered from 1 to 550 using a number stamping machine.

As an incentive for filling out the survey a drawing was going to be held for iPod Shuffles and iTunes gift cards with the number varying based on the class size. In order to ensure a fair drawing, a raffle ticket system was decided on. Each of the 550 surveys had a raffle ticket attached to it.

With 550 completed and numbered surveys and 550 half sheet instruction sheets, the final survey was ready to be distributed to students. At the beginning of the class period, instructions were doled out to the students. They consisted of who the researcher was, what he was working on, why he was doing it, and how the researcher thought they could help him. After the instructions were given the surveys were passed out and the students were allowed to complete them. No time limit was specifically set, instead, the researcher gave them the time they needed to finish the survey. The survey took people anywhere from five minutes to 15 minutes to complete.

Once all the respondents were done, the surveys were collected and the raffle tickets were entered into the drawing. In order to ensure that the students could not enter blank sheets, they turned in the full survey with half of the raffle ticket still on the sheet. This

allowed the researcher to very briefly flip through the surveys and make sure they were completed before entering the raffle ticket into the drawing. After all eligible tickets were entered, the drawing began. To continue with the idea of a fair drawing, students were randomly selected (and in some classes the professor drew the first ticket) to draw out and read the ticket number for the rest of the class. At the end of the drawing, the students were thanked for their time, the winners were congratulated, and the researcher promptly exited the classroom. This method was used in three different classrooms. One classroom had a different method, however. Dr. Tiwana took surveys to one of his classes, told his students what the survey was for, and then passed out the survey. They were asked to take the surveys home, complete them, and return them during the next class period.

Even though the surveys were individually numbered they needed to be tracked from class to class. Once a class was completed, the surveys from that class were given one of four different stamps to represent the four different classes. The stamp was placed on the first side in the bottom middle of the survey. To ensure that the surveys were not tampered with, they were locked in a faculty member's office with one exception. 100 surveys were removed from his office and transported to the researcher’s apartment for analysis for four days. The surveys were not removed from my apartment until they were transported back to Iowa State University.

3.6 Survey Sample and Data Collection

Iowa State University houses many different classes. Since a wide variety of majors were targeted, an large core business class was the most logical place to distribute the survey. The basic tenant of any business major is a large core undergraduate business class. This is a

class of typically greater than 100 people. A random sample was determined through these classes, and they were contacted in class where they were given a copy of the survey and a copy of the instruction sheet.

The respondents were first given verbal instructions and background on the survey, and then asked to read the instruction sheet. Once they finished the instruction sheet they could move onto the survey, where they were given 12 different peer-to-peer network profiles which they were asked to evaluate by rating two different dependent variables per profile. At the tail end of the survey, after the 12 peer-to-peer network profiles they answered, the respondents were asked to respond to demographic questions as well as providing their confidence level in relating the 12 peer-to-peer network profiles.

Four different classes were surveyed. The first class had a size of 60, the second had a size of 260, the third had a size of 140, and the forth had a size of 39. In total, the survey was given out to 499 people. In the first class, 53 people responded for a response rate of 88.33%. For the second class, 195 people responded representing a response rate of 75%. The third class had 108 respondents for a response rate of 77.14%, and the forth class had 24 respondents for a response rate of 61.54%. Overall, 380 surveys were responded to for an overall response rate of 76.15%. The numbers can be viewed in the figure below.

Class #	Class Size	# Respondents	Response Rate	% Response
1	60	53	.8833	88.33%
2	260	195	.75	75%
3	140	108	.7714	77.14%
4	39	24	.6154	61.54%
Overall	499	380	.7615	76.15%

Table 1-3 – Depicts the 4 classes chosen for survey distribution. The table shows the class size, the number of respondents, and the response rate in both decimal and percent.

3.7 Control Variables

8 control variables were established account for rival explanations in perceptions of peer-to-peer network content information assurance: (1) the level of confidence of the respondent in their evaluations, (2) the approximate number of hours per day that the respondent uses the internet, (3) how often the respondent uses peer-to-peer networks, (4) how long the respondent has been using peer-to-peer networks, (5) the respondents major, (6) respondent age, (7) respondents school year classification, (8) the respondents gender. The confidence level of the respondent was a single item assessment in order to determine the confidence degree of the respondent's assessments of the conjoint profiles. This variable controls for the respondents individual differences in their own assurance of the conjoin profile assessments.

Respondent Demographics and Descriptive Statistics

Multiple demographics questions were given for the respondents to answer. Of the respondent sample subset, 14 were MIS majors, 79 were non MIS majors, and 2 were dual majors. Since there were 100 surveys, 14% were MIS majors, 79% were non MIS majors, and 2% were dual majors. Three respondents were freshman, 41 were sophomores, 33 were juniors, and 14 were seniors. This translates into 3% of respondents being freshman, 41% being sophomores, 33% being juniors, and 14% being seniors. The third demographic question was in reference to the respondent's gender. 71 respondents answered that they were male, and 29 answered that they were female. This concludes that 71% of the respondents were men, while 29% were women. Respondents were asked to answer how confident they were in their responses to the conjoint profiles. In total, 93 people answered this question with seven abstaining. The mean for the respondent's confidence was 6.92, with a standard

deviation of 2.157. On average that means people were about 63% confident in their responses.

The next question had a response rate of 94 people with six abstaining. The respondents spent an average of 2-7 hours per day on the Internet. Since this question was broken down into four different answers and the mean answer was 2.5, that means that “2-4” and “5-7” were in between the mean. The standard deviation for this question was .726.

The next question had a response of 93 with seven respondents abstaining. The question was how often the respondent used peer-to-peer-to-peer networks, with the mean answer being 2.10 which corresponds to the answer of “occasionally”.

The final question had 93 responses with seven respondents abstaining. This question carried a standard deviation of .657.

Finally, the respondents were asked how long they had been using peer-to-peer networks. The mean answer was 2.97, which corresponds closely to the third answer, which was 3-4 years. The standard deviation for this question was 1.0.

Chapter 4 Analysis

4.1 Initial Analysis

4.1.1 Initial Data Entry

In total, 380 surveys were returned completed between the four classes where surveys were distributed. 380 surveys is far too large a sample to use for analysis purposes. As such, a smaller pool would need to be selected. A sample size of 100 was deemed to be a large enough sample that conclusions could be drawn. Since the third section had the closest number of respondents to 100, the third section was chosen to have a subset taken. 100 surveys were randomly taken from the section 3 respondents to be analyzed.

The reason that 100 respondents were chosen was to account for variance. The potential differences could be something as simple as more respondents in the morning class having more experience with peer-to-peer networks. So that the differences between the different classes were not considered, the 100 subset was chosen. What this means is that other class variances are not being considered for this study as they were not tested.

An initial table needed to be constructed to house the answers from the surveys. An Excel Spreadsheet was created. The spreadsheet housed 38 different columns and 100 rows. Each row contained the information for one survey. The columns contained the respondent's answers to each of the 24 dependent variables as well as their responses to the control variables. When a respondent did not answer a question no value was inserted making that entry null.

Some columns were coded a specific way to make information easier to track. The respondent's answers to the variable dependents, one per column, were a numerical value from 1 to 9. The overall confidence rating was a numerical value from 1 to 11, approximate

hours per day spent on the Internet was a numerical value from 1 to 4 to correspond to the 4 choices, how often they use peer-to-peer networks was a numerical value from 1 to 3 to correspond to the 3 choices, how long they have used peer-to-peer networks was a numerical value from 1 to 4 to correspond to the four choices, their major was a numerical value from 1 to 3 to correspond to the three choices, their status(i.e. Freshman, Sophomore, etc.) was a value from 1 to 4 corresponding to the four choices, their age was a numerical value from 1 to 6 corresponding to the six choices, and their gender was either a 0 or a 1, where 1 was used to identify females and 0 was used to identify males.

4.1.2 Data transformation

Profile#	Atttrib1 CntrbRep	Atttrib2 RateCont	Atttrib3 ContIden	Atttrib4 ContUse	Atttrib5 FileQual	Atttrib6 FileRate	Atttrib7 FileUse	Atttrib8 P2PTrust	Atttrib9 EaseOfUse
1	high	low	low	high	high	high	low	low	low
2	high	low	low	low	low	high	high	high	low
3	low	high	low	high	low	low	low	high	low
4	low	high	low	low	high	high	low	low	high
5	low	high	high	high	low	high	high	low	low
6	high	low	high	high	low	low	low	low	high
7	high	high	low	low	low	low	high	low	high
8	high	high	high	low	high	low	low	high	low
9	low	low	low	high	high	low	high	high	high
10	high	high	high	high	high	high	high	high	high
11	low	low	high	low	low	high	low	high	high
12	low	low	high	low	high	low	high	low	low

Table 1-4 – Shows the 12 different conjoint profiles with each of their 9 attributes. Each attribute has a “high” or a “low” to show if a peer to peer network uses a characteristic a lot or a little.

Once the data was entered into its initial form, it needed to be transformed into something that the analysis software could comprehend. Transformation of the original data consists of turning 100 rows into 1200 rows with corresponding information in each row. The first step taken was to make another Excel sheet. This sheet consisted of all 12 conjoint profiles in 12 different rows. Each row was a different profile consisting of all nine attributes

where all the attributes were labeled with either “high” or low.” The attributes were also color coded with “low’ being red and “high” being green. The end result was a sheet consisting of 12 rows with nine columns per row depicting 1 of 12 different conjoint profiles.

Once this sheet was constructed, the transformation could begin. Transforming the data meant first copying the color coded data into the transformation 100 different times to represent the 100 different surveys. This left the transformation with 1200 rows. The survey number was coded into the first column while the second column housed which conjoint profile data was input for which was a number from 1 to 12. Columns 3 through 11 were high and low values associated with a particular conjoint profile, and columns 12 and 13 were the respondents’ answers to that particular profile. Columns 14 through 26 were the demographic data taken from the back of each survey. For one set of conjoint profiles (all 12), the demographic data was inserted into the row making it the same for all 12 rows. This process was repeated for all 100 surveys in order for the data transformation to be complete.

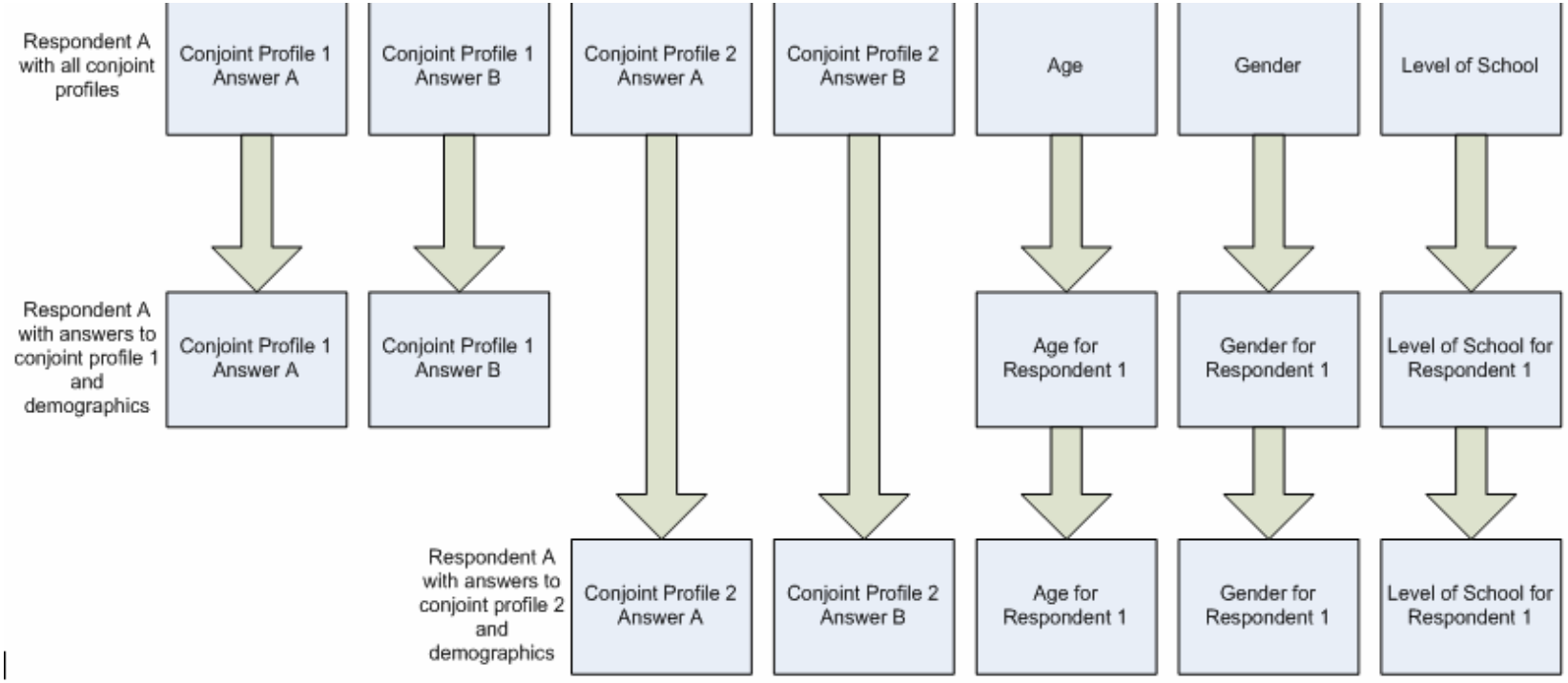


Figure 1-1 – Transformation of the survey data. Consisted of taking each individual survey, 100 in all, and breaking them into 12 different rows, for a total of 1200 rows. Each row represented one set of conjoint profiles and all of the demographic data.

It was extremely important that this step be done with the utmost care. 1200 different rows mean a lot of chances to incorrectly input data. Should something be off, such as a missing profile or data input into an incorrect slot, the conjoint algorithm would skew the results making them dreadfully inaccurate.

4.2 Testing

CONTRIBUTOR reputation	Number of positive ratings of this contributor by other users of this P2P network.
# of ratings for CONTRIBUTOR	Total number of other users that have rated this contributor (both positively and negatively).
CONTRIBUTOR identifiability	The extent to which contributors' identity can be verified in this P2P network (e.g., by requiring registration, valid email addresses, etc.).
Contributor reputation system usage	How extensively a contributor reputation/ rating system is used in this P2P network (e.g., users regularly leave comments and ratings on each other).
FILE quality rating	Number of positive ratings by previous downloaders of this file.
# of ratings for this FILE	Total number of previous downloaders that have rated this file.
File rating system usage	How extensively a file rating system is used in this P2P network (e.g., detailed comments left by previous downloaders).
P2P network trustworthiness	How much confidence you have that this P2P network is not flooded with fake/ malicious files (e.g., containing spyware or viruses).
Ease of use of ratings	How easy it is to use the file and contributor rating system to make download decisions.

Table 1-5 - Description of each characteristic in the search profile tables.

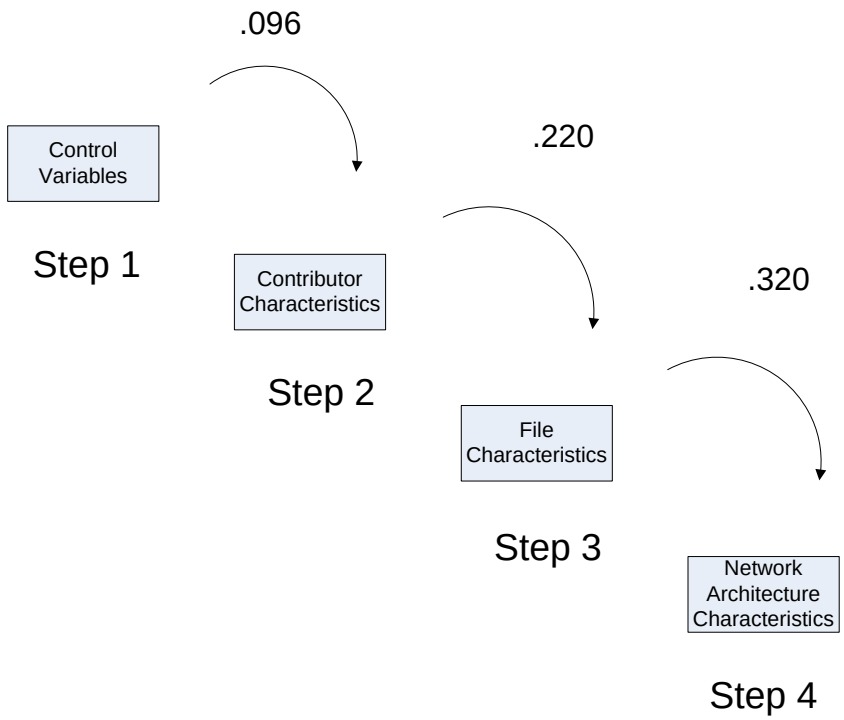


Figure 1-2 - This shows the step by step progression of variables added to the conjoint analyses

For the conjoint analyses, the association of the dependent and independent variables have two important measurements. First, the regression co-efficient, β, represents the relationship of the dependent and the independent variable. This tells how closely the two variables are correlated. The second measurement is the Z-statistic which tells the statistical significance of the correlation.

In order to test the data, a four-step hierarchical regression model was employed. The four steps were added incrementally, and those steps were: 1. control variables, 2. contributor characteristics, 3. file characteristics, 4. network architecture characteristics. These variables get added in an incremental fashion in order to help explain the respondent's decisions, as well as to help explain the contributory nature of each set of variables. No subgroups were required for this analysis. It was verified empirically that using these steps had no influence on the results.

4.3 Statistical Importance of Predictor Variables

Importance	Information Assurance Mechanism	β	t-Statistic
1	FileQual	.328***	13.138
2	P2PTrust	.246***	9.844
3	CntrbRep	.218***	8.733
4	ContIden	.138***	5.532
5	FileUse	.129***	5.162
6	RateCont	.127***	5.076
7	FileRate	.125***	5.013
8	ContUse	.085**	3.405
9	EaseOfUse	.079**	3.184

Table 1-6 – Breakdown of the 9 characteristics in order of their importance, along with their beta value and their t-statistic value.

The relative importance of the nine information assurance quality control mechanisms in peer-to-peer networks can be viewed in the table above. Based on the research, the file quality, or number of positive ratings by previous downloaders of this file, is the most important mechanism of the nine. The file quality had a beta value of .328 and a t-statistic of 13.138. The second most important mechanism was peer-to-peer network trust, which was how confident the user was in the network not being flooded with malicious files. The other seven mechanisms come in the following order: contributor reputation, contributor identity, file rating usage on the network, the rating of the contributor, the rating of the file, the contributor usage, and coming in last was the ease of use of the network. The top 7 mechanisms all had statistical confidence levels of .1%, and the bottom 2 had statistical confidence levels of 1%.

4.3.1 File Quality Characteristic

The most important characteristic based on the results of the study was the “file quality rating.” The file quality rating is the number of positive ratings by previous

downloaders of this file. What this means is that the more individuals that have downloaded this file and given it a positive rather than negative rating, the higher the overall rating. This is important because it shows an overall rating based on what your peers think of the quality and not some arbitrary meaningless number. For example: User A has a file called “Popular Song.mp3,” and user B also has a file called “Popular Song.mp3.” If user A’s file has a file rating of 6,909,809, that means over 6 million people have given user A’s file a positive rating. If user B’s file has a rating of 3,490, that means that three thousand people have given user B’s file a positive rating. According to the data, the respondents would be more likely to chose file A than file B.

The file quality had a beta value of .328 and a t-statistic of 13.138. Since the t statistic value is higher than 3.42, it means that this has a confidence level of .1%. The file quality rating characteristic positively influenced peer-to-peer network information assurance.

4.3.2 Peer-to-peer Network Trust Characteristic

The second most influential characteristic based on the results of the survey was peer-to-peer network trust. Peer-to-peer network trust can be defined as how much confidence the user has that a particular peer-to-peer network is not flooded with fake and/or malicious files. This means that a higher trust rating will occur if the user feels that a lot of files on this network are non corrupted and non infected files. However, this is merely perceived as a user cannot obviously know how many files are actually infected. For example: two networks exist, one with 1,000,000 files and one with 10,000 files. User A uses both of these networks. User A also knows that on network 1 there are 100,000 corrupt files, and on network 2 there

are 10 corrupt files. For network A, a total of 1 in 10 files are corrupted, whereas on network B a total of 1 in 1000 files are corrupted. User A will have a higher peer-to-peer network trust for network A than network B.

The peer-to-peer network trust characteristic had a beta value of .246 and a t statistic of 9.844. Since the t statistic was above 3.46, this characteristic has a .1% confidence level. This characteristic had a positive influence on peer-to-peer network information assurance, just not as much as the file quality rating.

4.3.3 Contributor Reputation Characteristic

The contributor reputation characteristic came in at the third most influential of the nine. The contributor reputation can be defined as the rating of this contributor by other users of this particular peer-to-peer network. This is a lot like the file rating characteristic except it relates to the user itself. For example, user A and user B are users of the same peer-to-peer network. User A has a high contributor reputation, 6734, whereas User B has a very low contributor reputation of 12. Users would be more likely to select user A to download from if both users are offering a file. To put this into a real world example, the contributor reputation characteristic would be EBay's feedback system. The higher their peers rate them, the higher their feedback is thus the higher their reputation.

The contributor reputation characteristic has a beta value of .218 and a t statistic of 8.733. Since the t statistic has a value higher than 3.46, it means that the confidence level is .1%. This characteristic has a positive influence on peer-to-peer network information assurance.

4.3.4 Content Identity

The fourth most influential characteristic was the contributor identifiability. The contributor identifiability is the extent to which the contributor's identity can be verified in this peer-to-peer network. This includes such aspects as the requiring of registration, providing valid e-mail addresses, or even age verification. For example, Network A has a fairly robust identity system requiring users to provide valid e-mail addresses, their age, date of birth, and proof that they are enrolled in the college they claim they are in. Users would be more inclined to trust a network such as this as opposed to a network that offers no identity verification processes.

The content identity characteristic has a beta value of .138 and a t statistic of 5.532. The t statistic is over the 3.46 threshold meaning this characteristic has a .1% confidence level. There is a positive influence on peer-to-peer network assurance for this specific characteristic.

4.3.5 File Rating System Usage

The fifth most influential characteristic is the file rating system usage characteristic. The file rating system usage is how extensively a file rating system is used in this peer-to-peer network. This can include things such as detailed comments left by previous downloaders, rating of other files, md5 hash check sums, etc.

The beta for this characteristic is .129 and the t statistic is 5.162. The t statistic is once again over the threshold of 3.46 meaning that the confidence level is .1%. There is a positive influence on peer-to-peer network assurance for this characteristic.

4.3.6 Number of Ratings for the Contributor

The sixth most influential characteristic is the number of ratings for the contributor of the content. This number represents the number of previous downloaders of the contributor's content that have rated the contributor. This could mean that the individuals have given either a positive, negative, or neutral review of the contributor. An example is the EBay feedback system. In their system, a number is placed in parenthesis next to the user ID: an example would be averagejoe(16). This means that averagejoe has been given a rating of 16 which is calculated by taking the total number of positives and subtracting the total number of negatives. The higher the number the better the user's reputation.

The characteristic had a beta value of .127 with a t statistic of 5.076. The t statistic is over the threshold of 3.46 which means the confidence level for this characteristic is .1%. There is a positive influence on peer-to-peer network information assurance.

4.3.7 Number of Ratings for a File

The seventh most influential characteristic is the number of ratings for the specific file. This characteristic can be defined as how many of the previous downloaders have rated this file. These ratings include positive, negative, and neutral ratings. For example, a file may

have had 73 positive ratings, 16 negative ratings, and 11 neutral ratings, giving it an overall file rating of 100.

This characteristic had a beta value of .125 and a t statistic of 5.013. The t statistic has a value of over .346 making this the last to have a confidence level of .1%. There is a positive influence on peer-to-peer network information assurance.

4.3.8 Contributor Reputation System Usage

The second to least important characteristic of the nine was the contributor reputation system usage characteristic. This characteristic measures how extensively a contributor reputation or rating system is used in this specific peer-to-peer network. When looking at the network this would be a measure of how many people are actively participating in the rating/reputation system or how wide spread the system is. For example, a network may have a reputation system that is being used be 90% of the population of that network, and a different network may have a reputation system that is being used by 10% of the population. For this example, the contributor reputation system usage is higher for the first network than the second.

The beta for this characteristic was .085 with a t statistic of 3.405. Since the t statistic is below 3.46, the confidence level drops from .1% down to 1%. There is a positive influence on peer-to-peer network assurance for this characteristic.

4.3.9 Ease of Use

The least important characteristic was related to the ease of use in the rating system. This characteristic is defined as how easy it is to use the file and contributor rating systems to make download decisions. The ease of use really comes down to how the system is easy and intuitive to use, or is it bulky and time consuming to leave ratings? For example, a network has a contributor rating system that requires you to login, find the file, and when you leave a rating it requires to you enter some personal information about downloading habits. A different network has a contributor rating system that simply requires you to login, click positive, negative, or neutral, and then submit it. The second network has a much more simplistic rating system whereas the first one has a bulky rating system.

This final characteristic has a beta value of .079 and a t statistic of 3.184. The t statistic is below 3.42 so the confidence level is 1%. There is a positive influence on peer-to-peer network assurance for this characteristic.

Chapter 5 Discussion

5.1 Limitations

5.1.1 Assumptions

This study incorporates a small number of premises. First, this study was designed to be given to end users of internet peer-to-peer networks. The study was also designed to question the end user on which of the nine listed information assurance mechanisms affect whether or not a user will download a file. It is assumed that the participants will answer the questions voluntarily and honestly. Lastly, it is assumed that the survey participants will take the time to read through the instruction sheet and only respond to the questions that they understand.

5.1.2 Possible Problems

When the survey was being refined only one graduate student was given the survey. The control variables consist of 7 different questions one of which asks the students status. This question has 4 different choices for answers: Freshman, Sophomore, Junior, or Senior. Graduate student was not a choice that was written down, simply because the researcher did not think about it. However, once the surveys were passed out and the answers were collected the question came up. What if someone in one of those classes was a graduate student? The researcher had assumed that no graduate students would be in an undergraduate level core business class because it is an undergraduate class, and the researcher did not know that any graduate level student might have to take it. However, it was pointed out that graduate students who transfer to the business college are sometimes required to take

undergraduate courses or sometimes voluntarily take them. A fifth status, graduate, should have been added into the survey from the beginning, but the researcher do not think that leaving this out of the survey will have the outcome of skewing the results.

There is a possibility that the survey itself could have tainted or skewed results. It is not overly unlikely to think that the respondents either lied about their answers, randomly chose answers, or some other nefarious thing. The results of the survey really do depend upon the trustworthiness of the respondents. One possible way to check for these errors would be to analyze every single respondent's survey. The likelihood that a portion of the people from a sample of 100 lying would be greater than the likelihood of 400 respondents lying. However, when using the 100 survey sample, some of the respondents did not answer some questions. Some, for whatever reason, did not answer a few of the conjoint profiles or a few of the demographic questions. These profiles were not discarded and their answers were entered as null values. On a larger scale the number of respondents that chose this path would obviously be higher, but it could also be claimed that the results take these into account, which means no tainted or skewed results.

An area that obviously has the potential to go wrong would be researcher error. It is not unlikely to think that the researcher made one or more errors when imputing the information from the collected survey data into the computer. The human error factor is still there, and there would have been multiple places for an error to have been made. However, the 100 surveys used for analyses were kept separate from the rest of the surveys, and kept indefinitely if such an error were to be made. It should also be noted that if the imputer

entered a 1 where a 0 was supposed to go, or something equally as minute, the overall results of the survey would have changed little if even at all.

With the exception of simply leaving off graduate students as a demographic, nothing really went wrong here. There were no problems analyzing the data, collecting the data, or the creating the survey and the students had no questions during the survey. To me this signifies that the collection process and analyses process went really well.

5.2 Future Research

Ideally there is a direction that the researcher would like this research to go. The researcher would like to deploy the survey at other schools throughout the nation to get a healthy sample from all over. The survey could be deployed at colleges big and small throughout the Midwest, west coast, east coast, and everywhere else. Once the survey has been deployed an analysis would be conducted based on region. The analysis would help determine, by region, the effect that the 9 listed information assurance mechanisms would have on the end users downloading habits. The researcher would do this by region in order to determine if there was a regional bias toward some or all of the mechanisms or if the results were fairly common throughout the entire country.

If the analyses showed that the results were fairly common throughout the country, this research would go one step further. The researcher would like to actually implement those 9 mechanisms based on order of importance into 4 peer to peer networks, based on the 4 different categories of peer to peer networks identified earlier. These networks would be deployed in the hopes that end users would see the mechanisms in place and that would

convince them of 2 things. The first is to actually use and maintain use of said network and the second is the reassurance, based on the mechanisms, that the files they are downloading are most likely what they claim to be. This would perhaps involve other peer to peer networking companies (Google, YouTube, Skype, EBay etc). With the help of one or more of these companies the networks could be deployed and maintained on a much larger basis than if an individual, such as myself, were tasked with deployment and maintenance.

Another possible future research field would be analysis by network type. Since there are different network types and these were not taken into explicit consideration, it would be of use to determine if there is a variance between networks. This stems from the fact from that the new incarnation of Wikipedia is based on the contributor mechanisms. Since the new Wikipedia is being based on contributor mechanisms it stands to reason that there may be a variance of mechanism importance across peer-to-peer networking types.

It is also the hope of the author that other researchers and companies see peer to peer networking for what it is: an emerging technology that, when properly maintained and thought out, could benefit users of the internet in a significant way.

Chapter 6 Conclusions

This study attempted to ascertain how end user perceptions were affected end user perceptions of downloading in a peer to peer network environment. In total 9 different hypotheses were tested. The results were overwhelmingly positive. The 9 hypothesis all showed a positive influence on peer to peer network information assurance. What this means is that of the 9 hypothesis, each had an effect on the perceptions of the end user when they were making a decision on whether or not to download a file. The most influential mechanism was file quality rating characteristic. This means that the number of previous downloaders of a file that gave the file a positive rating had a larger affect on the users downloading perception than any of the other 8 characteristics. The least influential was the ease of use characteristic. This means that how easy the rating or reputation system is to use will have the least effect on the end users downloading perceptions out of all the characteristics.

In closing the results of this study were very positive and promising. There is a significant amount of follow up research that can be done using what has already been presented. It is my hope that this thesis paper will allow for expansion into more research and a greater understanding of the full potential that peer to peer networking has to offer the internet community as a whole.

Table 1-1 – 100 Peer-to-Peer Networks

Network	URL	Content Type	Major Quality Control Mechanisms	Age of Network	Size	Major Problems
Video						
Google Video	video.google.com	Video	User ratings, Top 100, Other related search fields, top few videos per category on the front page, number of views, date posted, number of ratings, comments, other videos from same user	< 1 year	unknown	Copyright Infringement, extreme bandwidth usage/costs, misleading video titles, information integrity
YouTube	youtube.com	Video	user ratings, number of ratings, listing of related clips, user comments	<= 1 year	unknown	Large bandwidth, no real revenue structure No way to rate channels, almost all channels are listed in foreign languages, some random file types for broadcasting, copyright infringement, information integrity photos you may not want up there can be listed, no way to have photos you don't want posted removed,
Peercast	peercast.org/	radio and video	station name listed, the number of connections listed, the rate at which it is going is listed, the file type is listed, the total number of channels listed, the total number of listeners listed, the total relays listed, ability to search by genre/speed/type/status,	4 years	unknown	
Flickr	flickr.com	Photo	list of popular search queries, user comments, listing of how many pictures a user has posted	approx 1 year	unknown	

iTunes	apple.com/itunes	Audio/Video/Podcast	User ratings, user comments, iTunes allows users to share out their music so others can stream it without downloading it, anyone can create and share a podcast	approx 5 years	unknown	restrictive DRM, pay to download, bandwidth from streaming, comments are often made by immature people with grudges against a song/artist/podcast random file naming by users, copyright infringement, mostly just podcasts, rip off of iTunes without the nice client application
Odeo	odeo.com	Audio	total subscriber numbers for the podcasts, user comments, total number of plays, when video was posted,	approx 2 years	unknown	potential problems seen is a user created movie could be absolutely horrid to watch, could have problems, not be coherent etc
A Swarm of Angels	aswarmofangels.com	User created movie	no DRM, ability to burn it to DVD, watch it on an ipod or PSP, ability to freely remix it, licensed under creative commons	< 1 year	< 1000	No user ratings, no user comments, no number of views, no top 100 or top 10, no date posted
Photobucket	photobucket.com	Photo	total user pics, ability to see all user pics, ability to get file size, file data, resolution, can embed videos into other web pages	3 years	unknown	no way to view other users pictures, no pictures on the front page, no search function, copyright infringement, bandwidth costs, no ratings, slow site, no hotlinking
imageshack	imageshack.us	Photo	no account is required,	3 years	unknown	

CyberskyTV	tvoon.de/ctv	Television	no DRM limitations, over 10000 channels, all done through peer-to-peer so no intermediate to charge,	unknown	10000 channels	no ability to pick which peers you connect to, requires a lot of bandwidth, must install software, no way to rate stations, no way to rate users
CoolStreaming	all-streaming-media.com/peer-to-peer-TV/	Television	The more users the faster it goes, total number of users per channel is listed, channel names are listed, quality of the station in % listed	unknown	unknown	must be uploading for at least 20 minutes before you can start to download which makes trying to watch live TV shows very difficult, copyright infringement, currently only offering Chinese and a few Italian shows, no user ratings, no web interface, no way to rate pictures,
Hello	hello.com	Pictures	works with picasa and blogger, able to view all of users pictures, 128 bit AES encryption,	2 years	unknown	no total views of pictures,
All Peers	allpeers.com/index	Pictures, audio, video, websites	ability to share only selected files, ability to share only with those whom you deem worthy to share with, since you can choose who to share with you can have the security of not getting caught file sharing, done through your FireFox web browser,	unknown	unknown	no way to rate files, information integrity, copyright infringement, not many people use this, cannot share files with the general public
Buzznet	buzznet.com/video	Video	featured videos, when posted, how many views, comments, number of comments listed, user pages, user page comments, look at all user videos	unknown	unknown	person pages are cumbersome and difficult to read, the rating system is simply a number from 1 to infinity, a

Viral Video	viralvideo.cleevver.com	Video	featured videos, when posted, how many views, comments, number of comments listed, look at all user videos, can score the video	> 1 year	unknown	lot of videos are just hotlinked to myspace and YouTube, bandwidth costs, copyright infringement Unknown who posted the video, copyright infringement, the "score" is a simple number, bandwidth costs, ads right next to the video
Clipshack	clipshack.com	Video	popular topics, latest clips, feature clips, who uploaded it, how long ago it was uploaded, number of views, what topics/tags the video falls under	unknown	unknown	no user ratings, copyright infringement, bandwidth costs, competing against YouTube, information integrity, users not putting in comments but random gibberish
Myspace Video	vids.myspace.com	Video	who uploaded, date uploaded, number of views, front page features featured/most popular/recent uploads, video ratings, view all videos by a user, synopsis of video, top 100 , user ratings on front page	3 years	unknown	Clicking on the person takes you to their obnoxious myspace pages, bandwidth costs, copyright infringement, no user/identity management,
KoffeePhoto	koffeephoto.com/en/index.php	Pictures	all pictures are stored on your local machine, can set total size limit for pictures, can setup to use real name, total number of users connected listed, total number of pictures listed, total size of all pictures listed, total number of albums, average user availability	unknown	unknown	copyright infringement, no way to know how long someone has been a member for, no way to rate photos

Yahoo! Video	video.yahoo.com	Video	number of views, video rating, number of raters, user reviews, when video was posted, video source, featured videos per day, video tags for easier relation, ability to flag as offensive, view all videos posted by user	unknown	unknown	potential for copyright infringement, complementation against other more popular video sites, trying to build a user base, no top 10 or top 100, some videos aren't housed at yahoo but linked to other sites, potential for serious misuse of non yahoo hosted videos, bandwidth cost
iFilm user video	ifilm.com/uservideo	Video	who uploaded, date uploaded, number of views, front page features featured/most popular/recent uploads, video ratings, view all videos by a user, synopsis of video, top 100	> 4 years	unknown	no ratings on the front page, a lot of videos aren't even rated, copyright infringement, bandwidth cost
eyespot	eyespot.com	Video	File author, new additions, recent postings, when file was posted, most popular, ability to mix files and flag them as mixes, can view all things posted by a user	unknown	unknown	No user join date, no user rating, no user comments, no file ratings, no file comments, copyright infringement, bandwidth cost
Grouper	grouper.com	Video	file author, when file was uploaded, how many views the file has, the channels it falls under, file tags, number of user profile views, number of users video views, number of videos shared, user comments, most viewed videos, highest rated videos	< 1 year	unknown	bandwidth cost, copyright infringement,

JumpCut	jumpcut.com	Video	Who submitted video, when it was submitted, number of views, rating from users, user comments, number of movies user is sharing, comments from other users	< 1 year	unknown	The rating system seems weak and unused, copyright infringement, bandwidth costs
Revver	one.revver.com	Video	Average rating, who uploaded, number of views, most watched is posted,	unknown	unknown	Money costs since users get paid for every time their video is viewed, average review doesn't tell how many people reviewed it, must login to view user information, bandwidth cost
Vimeo	vimeo.com	Video	When file was posted, who posted it, how many people "liked" it, when the member joined, how many clips the person has posted, how many they have liked, user comments	unknown	unknown	A lot of posted videos are more than 1 year old, copyright infringement, doesn't list the total number of views, bandwidth cost
vSocial	vsocial.com	Video	Who submitted video, when it was submitted, number of views, number of comments, user comments, identifying tags, rating out of x votes,	4 years	unknown	no user join date, no total number of views a user has, no way to see the rating of a video before clicking on it, a lot of comments are in different languages, bandwidth cost
CastPost	castpost.com	Video	When video was posted, who posted, how many times It was viewed, how many comments, users other posting with how many comments each has, file name listed	2 years	> 5000	copyright infringement, multiple languages, no user rating system, no user join date, no video rating system, bandwidth cost

Sharkle	sharkle.com	Video	who uploaded, date uploaded, total views, total comments, user rating out of x voters, last view, user comments, user join date, favorite videos, other uploaded videos by user w/ their ratings and total views	< 1 year	unknown	copyright infringement, no rating before you click on the movie, very few "top ranked" or "featured" videos have comments, no way to view users other comments, bandwidth cost
Blip TV	blip.tv	Video	when video was posted, who posted, what license the video has, what tags it has, rating out of 5 stars, ability to flag something as inappropriate	< 1 year	unknown	no number of views, no user join date, no user rating, does not tell what the number of raters were, user abuse of the report as inappropriate feature, bandwidth cost
Daily Motion	dailymotion.com	Video	featured videos, who submitted, when it was submitted, what language submitted in, the users most popular video, when user registered, average rating out of x votes, lists your personal rating of it	< 1 year	unknown	no user rating, many different languages making finding some videos difficult, bandwidth cost
Veoh	veoh.com	Video	number of views, rating system, who posted it, when it was posted, how, when user joined, users average ratings, how many videos they have published,	2 years	unknown	Massive copyright infringement, doesn't list ratings out of x people, cant view all of the users comments, doesn't tell the users rating out of x people, bandwidth cost
Guba	guba.com	Video	total videos in a category, rating out of 5 stars, who uploaded it, the number of views, the actual file name, what resolution its in, user comments, how many videos the user has uploaded,	8 years	unknown	Massive copyright infringement, does list ratings out of x users, doesn't have a user join date, no user ratings, no way to

						view all of a users comments, bandwidth cost
Vmix	vmix.com	Video	who uploaded the video, what its average score was, how many people scored it, the number of views, my score for it, user comments,	unknown	unknown	doesn't list when the user joined, no use rating, no way to view the users comments, copyright infringement, doesn't list rating until you click on the video, bandwidth cost
VideoSift	videosift.com	Video	who posted the video, when it was posted, what the videos rank is out of how many videos, how many votes for the video, number of user comments, star rating(in colors) to show how many videos user has promoted to the front page, user joined date, user comments on user profiles	approx 1 year	unknown	cannot vote negative for a video unless you are a certain rank member, a lot of videos linked directly from other sites, bandwidth costs, copyright infringement
Audio						
Kazaa	kazaa.com	Audio/Video/Software/Pornography	Total files a user is sharing,	Alive for approx. 4 years	unknown	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties

Bearshare	bearshare.com	Audio/Video/Software/Pornography	Can choose which peer to connect to, total user files listed, file ratings to show relevance, file sizes listed, relevant files listed by size to choose appropriate one.	live a few years	unknown	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties
Limewire	limewire.com	Audio/Video/Software/Pornography	file ratings, use of SHA-1 and tiger tree cryptographic hash functions	approx 6 years	2 million + users	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties
BitTorrent	bittorrent.com	Audio/Video/Software/Pornography	Total seeders and leechers for a file, file ratings, when file was posted, last time file was seeded, many different users to download from to increase speed, large amount of search engines with many files, group that released the torrent is listed, with the ability to see what other files that group has released	approx. 4 years	impossible to know	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties, no way to specify which peers to connect to and which ones not to, no viewable peer rating of those you are connected to
Napster	napster.com	Audio	New - music came from centralized Napster server, meaning you knew what you downloaded was legit, in the original there was a central server where files were stored offering better information integrity	Alive for approx. 4 years	26.4 million	Original - centralized server for content distribution got them in trouble for copyright infringement, information integrity once database was changed, New - music is only rented, and you cant own it, no ability for user generated content, competing against iTunes too late with

						too little to offer
AudioGalaxy	audiogalaxy.com	Audio	anonymity - you didn't know who you were downloading from, songs/files had ratings	alive for approx 2 years	unknown	Information integrity, no ability to select the fastest "peer" to download from
DC++	dcplusplus.sourceforge.net	Audio/Video/Software/Pornography	ability to search all of a users shared files, files are rated, shows full bit information to check for validity	multiple years	unknown	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties
IRC	mirc.com	Audio/Video/Software/Pornography	word of mouth, ability to know how many files a person is sharing as well as their network speed and download caps/queue numbers	approx 18 years	unknown	slow speeds, difficult to find files, download queues are excessive, information integrity
WinMX	zeropaid.com/winmx	Audio/Video/Software/Pornography	Can choose which peer to connect to, total user files listed, file ratings to show relevance, file sizes listed, relevant files listed by size to choose appropriate one.	Alive for approx. 4 years	unknown	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties
morpheus	morpheus.com	Audio/Video/Software/Pornography	file details and availability information, forums to allow user interaction, file ratings	approx 4 years	> 20000000	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties
grokster	grokster.com	Mostly audio	Can choose which peer to connect to, total user files listed, file ratings to show relevance, file sizes listed, relevant files listed by size to choose appropriate one.	approx 3 years	unknown	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties

gnutella	gnutella.com	Audio/Video/Software/Pornography	Can choose which peer to connect to, total user files listed, file ratings to show relevance, file sizes listed, relevant files listed by size to choose appropriate one.	approx 5 years	approx 2.2 million	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties
eMule	emule-project.net/home/perl/general.cgi?l=1	Audio/Video/Software/Pornography	Can choose which peer to connect to, total user files listed, file ratings to show relevance, file sizes listed, relevant files listed by size to choose appropriate one.	4 years	unknown	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties
eDonkey	edonkey.com	Audio/Video/Software/Pornography	Can choose which peer to connect to, total user files listed, file ratings to show relevance, file sizes listed, relevant files listed by size to choose appropriate one.	2 years	2000000	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties
shareaza	shareaza.com	Audio/Video/Software/Pornography	Can choose which peer to connect to, total user files listed, file ratings to show relevance, file sizes listed, relevant files listed by size to choose appropriate one.	unknown	unknown	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties
Soulseek	slsknet.org	audio	if multiple files are present, ability to compare, file size listed, relevance listed, what album the song comes from listed, other user files listed, total user files listed	unknown	unknown	Problems connecting to servers, problems with speed of peers, no way to designate which peer to connect to, information integrity, copyright infringement, search difficulties, virus/spyware issues
i2hub	i2hub.com	Audio/Video/Software/Pornography/homework	total files listed, ability to browse all of a users shared files, file sizes listed, connection speed of user listed, total download slots of user listed, file relevance	1.5 years	> 10000	Copyright Infringement, Information integrity, network speed, virii/spyware, search

						difficulties
Usenet	usenet.com	Audio/Video/Software/Pornography		unknown	unknown	
iMesh	imesh.com	Audio/Video/Software/Pornography	Can choose which peer to connect to, total user files listed, file ratings to show relevance, file sizes listed, relevant files listed by size to choose appropriate one.	unknown	> 100000	Copyright Infringement, Information integrity, network speed, virii/spyware, search difficulties
Soribada	soribada.com	Audio	now a fully legit pay service, information integrity problems would be kept to a minimum, downloading from other peers legally	6 years	unknown	In Korean so very difficult to use correct, original had copyright infringement, information integrity, network speed, virus/spyware, search difficulties
Software						
SourceForge	sourceforge.net	Opensource software	project of the month, who posted, date posted, submitting bugs, viewing bugs online, view users projects, project lists project admins, project comments, total number of bugs, total number of downloads	5 years	> 1 million	no way to rate the project, no user comments other than bugs
Skype	www,skype.com	peer-to-peer telephone	information about the person, advanced searches to find someone, user can be identified by many qualities so you know who your getting, user profiles, specify your own user comments	3 years	> 4000000	no rating system, no way to tell if who your talking to is the right person, no age/identification

						management,
Opensource Tutorials	opensourcetutorials.com	Opensource tutorials for coding languages	lists what user posted the tutorial, lists when it was posted, lists the users resources if they put them in, forums with number of threads, number of posts, who the thread author is, user level, when user joined, when their last activity was, number of user posts, ability to find all posts by a user, ability to find all threads by a user	3 years	> 500	other than the forums there is no way to rate an article on the website, very narrow well defined articles, articles do not appear to be updated often or have newer content, possible copyright violations
Project Gutenberg	gutenberg.org	Opensource books	top 100 books and authors, advanced searches, total number of books listed, total number of books downloaded each month, copyright status listed, date book was released	3 years	19000 books	No way of rating the book, no way to tell which user got it online, no user ratings,
Gutenberg distributed proofreaders	pgdb.net/c/default.php	Opensource book proofreaders	ability to only work on a chapter at a time, number of projects completed per month, number of active users broken down into periods, recent completed projects, number of projects completed, number in progress, number in proofreading,	6 years	> 10000	no way to tell who did what for the book, no rating system, no quality control, relies strictly on the users good judgment and honesty, no search function, no user comments
Freshmeat	freshmeat.net	open source projects	when posted, who posted, what license posted under, what changes were made, when user was created, all user projects, all user comments,	4 years	> 369000	no user rating, potential for copyright infringement
D-Zone	dzone.com	development projects	who posted, total votes, when posted, total views, when user joined, number of links posted, number of votes received, votes given to others, users get to vote what goes on the front page	unknown	unknown	no user rating, potential for copyright infringement

Deals						
Slickdeals	slickdeals.net	Deals forum	total threads, total posts, thread started, thread posts, thread views, last thread post, user rating, user posts, user join date, user reputation, thread ratings	unknown	> 150000	false information, expired deals, misleading threads/topics, exclusive member only deals posted
Techbargains	techbargains.net	Deals forum	total posts, total topics.total thread replies, total thread views, thread starter, last post,	approx 7 years	> 50000	false information, expired deals, misleading threads/topics, exclusive member only deals posted
FatWallet	fatwallet.com	Deals forum	thread status, thread ratings, total thread posts, total thread views, thread originator, age of thread, last post time, member status,	approx 7 years	unknown	no way to view a users profile, can go in and change votes at will, "flag size" is not a good rating system, kind of ambiguous, no ability to see how many other threads a user has started, no way to see what users other thread ratings were
EBay	ebay.com	Auction hub	user ratings, feedback, quality control done by eBay, dispute console, numbers rating for transactions completed, ability to see total ratings from buyers and sellers separately	approx 11 years	> 100000000	fraud/false auctions, information integrity, dishonest sellers/buyers

Wikipedia	en.wikipedia.com	Wiki/information hub	ability for anyone to edit any content, total number of topics, total number of edits, what the edits were, user profiles, difference between edits listed	approx 5 years	approx 50000	Ability for anyone to edit any content
Digg	digg.com	News	"digg up and digg down" stories, overall number of diggs, user ratings, user comments	Approx 2 years	> 500000	potential for small group of users to "dominate" the front page stories
Newsvine	newsvine.com	News	list of users friends, list of the articles and seeds by the person, user comment, story ratings, number of articles posted, number of links seeded, number of comments per story, users submitted stories on the front page	approx 1 year	> 10000	rude and mean users, hard to figure out personal "blog" section, no way to tell if an article has already been posted that you are going to post, users seemingly rate based on the current trend (i.e., rate you up if your ranked high or down if your ranked low)
Dealnews	forums.dealnews.com	Deal forums	thread started by, last post, number of posts, total number of posts by a user, user registration date, user private messages, total numbers if threads and posts per forums, last forum post	approx 10 years	unknown	no thread rating, information integrity, exclusive deals, horrible site layout, lack of age/identity management, ability to post fake myspace accounts, owned by NewsCorp, bandwidth costs, information integrity
Myspace	myspace.com	Personal interactions	Ability to take your profile private, monitored by myspace for inappropriate content, user comments, ability to block members, must be approved by owner of page in order to post on it	approx. 2 years	100000000	

Facebook	facebook.com	Personal interactions	to register as a student must have an edu account, ability to search based on many criteria, lists persons "real" name, user comments, user personal information, list of user friends, lists users groups	approx 3 years	9 Million	horrible site layout, lack of age/identity management, ability to post fake accounts, recently went public so anyone can make an account, bandwidth costs, information integrity
Friendster	friendster.com	Personal interactions	users last login, users join date, total friends, total pictures,	approx 4 years	31 Million	lack of age/identity management, bandwidth costs, information integrity, ability to post fake accounts,
Del.icio.us	del.icio.us	Personal bookmark sharing	number of users that saved the bookmark, top bookmarks, keywords, total number of user bookmarks, how long user has been posting, how much time between posts	approx 2 years	unknown	cumbersome front page, no way to rate a link,
Got Apex	gotapex.com/got-deals	Deal forums	region specific deals, total replies, total views, user "ranking", user join date, total user posts, view all posts by user, view all threads started by user, total users viewing a forum, number of registered and unregistered persons online, users personal profiles	approx 6 years	> 20000	no thread rating viewable in the forum, a lot of users do not put the expiration date in the thread or title, some offers are not for general public, bad links, expired deals etc etc
Craigslist	craigslist.org	Classifieds	ability to email the seller, total number of "things" in each category, ability to view by state, or by topic, discussion forums, date the listing was posted,	approx 10 years	unknown	no product rating system, no user rating system, no way to tell how long a user has been registered
Anandtech	anandtech.com	Deal forums	total number of topics, thread replies, thread views, user status, number of user posts, user joined date, last post date, average user posts per day	approx 6 years	> 150000	no thread rating viewable in the forum, a lot of users do not put the

Deal Catcher	dealcatcher.com/forums/forumid_21/tt.htm	Deal forums	total number of topics, total number of posts, total number of views, user ratings, user joined date, user status, user joined date, ability to see users most recent posts, ability to read users posts	approx 6 years	approx 50000	expiration date in the thread or title, some offers are not for general public, bad links, expired deals, cant view member profiles, cant view other threads started by member, cant view members other thread ratings no thread rating viewable in the forum, a lot of users do not put the expiration date in the thread or title, some offers are not for general public, bad links, expired deals etc etc
Deal of Days	forums.dealoffday.com	Deal forums	number viewing each category, numbers of threads, number of posts, number of views, total user computers, member status, number of referrals, ability to view threads started by user, ability to view all posts by user, user join date	approx 6 years	> 150000	some forums have only one poster, no thread rating viewable in the forum, a lot of users do not put in the expiration date, some offers are not for the general public, bad links, expire deals, cumbersome front page, custom user pages are difficult to view, must compete against the giant related sites, large bandwidth usage,
Bebo	bebo.com	Personal interactions	user profiles, number of page views, users last login, comments from other users, links to the users friends, ability to clock users, able to report abuse, skype integration	approx 1 year	unknown	

hi5	hi5networks.com	personal interactions	connection path from user to user, user comments, user last login date, ability to block the user, abuse reporting, total page views	unknown	> 50000000	a lot of pages in different languages, auto playing music on most pages, copyright infringement, ability to create fake accounts, no age/identification verification
Xanga	xanga.com	Personal interactions	user comments, "eprops," blocking of members, allowing only certain members to read your site, self ratings, community ratings, total comments, total eprops, original post date	unknown	unknown	imbedded audio/video files, pages in different languages, user created pages that are obnoxious and hard to read, competing against the bigger sites, coming up with the user base
Deal Database	dealdatabase.com/forums	Deal Forums	total threads, total posts, last post, total number viewing, thread replies, thread views, who posted the thread, member join date, total member posts, find all posts by member, find all threads started by member, member status/ranking,	approx 6 years	> 50000	no way to rate a thread, information integrity, no dates listed in thread title for deal expiration, no thread rating, information integrity, seems to be dominated by a few posters, the few posters could control what's seen and not seen, thread icons don't relate to the thread topic and are user selected
Flamingo World	flamingoworld.com/forum/ubbthreads.php	Deal Forums	total threads, total posts, suggestions forum, thread originator, thread views, thread replies, last post, registered date, member number, total posts, user title,	approx 8 years	> 30000	

Fishing for Deals	fishingfordeals.com/forums/postlist.php?Cat=&Board=catch	Deal Forums	"important phone numbers" thread, total views, total replies, thread originator, last post in thread, total user posts, user registration date, ability for users to reply and comment	5 years	> 15000	no thread rating, no dates listed in thread titles, thread icons have nothing to do with the thread, no user ratings, no user ratings
Big Big Savings	forums.bigbigsavings.com	Deal Forums	total forum statistics, forum viewers, number of threads, number of posts, last post, user join date, total user posts, total user posts per day, find all posts by user, find all threads by user,	6 years	> 34000	no thread ratings, no user ratings, thread icons don't relate to the thread, no way to compare users previous started threads or posts for accuracy
Bargain Share	bargainshare.com	Deal Forums	total threads, total replies, last reply, topic started listed, when thread was started, total replies, total views, last replier, user join date, total user posts, which forum they are most active in, if they are currently logged in, user status	4 years	> 50000	no viewing number listed, no way to view users other posts and other threads, no thread ratings, no user ratings that seem to mean anything
What's your deal	whatsyourdeal.com/forums	Deal Forums	total topics, total replies, total views, thread originator, last post time, total posts, join date, Rep Power, users title/rank, number of referrals	2 years	> 5000	no thread rating, no deal expiration date, forums seem to be dominated by a few users, thread icons have nothing to do with the thread, no user ratings
Daily Freebies	freesamplesite.com/ydf/	Deal Forums	total threads, total posts, number viewing a forum, thread originator, thread views, thread replies, last thread post, thread origination date, user join date, number of user posts,	6 years	> 20000	no thread rating, no deal expiration date, thread icons have nothing to do with the thread, no user ratings, no user ratings

DSL Reports	dslreports.com/forum/hotdeals	Deal Forums	last poster, last post time, replies and unique replies, account type, joined date, total posts, reviews submitted, last login, total number of topics, list of topics started within last 14 days, list of posts started within last 14 days, frequent posters listed at the side	7 years	> 80000	requires registration to view any thread, must pay to use advanced features, must pay for no advertising, must pay for special site features, no user ratings, no thread ratings,
SysOpt Forums	sysopt.com/forum/forumdisplay.php?f=14	Deal Forums	total viewing, total threads, total posts, last post, last post time, total user posts, user join date, last post by user, all posts by user, all threads started by user,	4 years	> 80000	no thread ratings, no member ratings, personal information could be inaccurate,
RedFlag Deals	redflagdeals.net/forums/forumdisplay.php?f=9	Deal Forums	thread ratings, thread replies, thread views, thread originator, last thread poster, category type listed sometimes, user posts, users community ranking, user join date, find all posts by user, find all threads by user, personal information can be entered'	6 years	> 58000	few threads actually have ratings, user ratings don't tell much about the user,
My What a Deal	mywhatadeal.com/index.php	Deal Forums	thread posts, thread views, total user posts, when user joined, user status, what forum user is most active in,	unknown	> 200	No way to view all of a users posts, no way to view other threads a user has started, no user ratings,
MyFeedz	myfeedz.com	Social News	how many related articles there are to it, where the article came from, article ratings, can view all "feeds" from a user, top feeders listed	approx 1 year	unknown	no user ratings, no telling which user posted it, no user profiles to view,
WobBlog	wobblog.com	Social News	number of "wobbles" a story gets, how many comments it has, top users section, when user registered, total links they have, total links that were published, total votes, total comments	> 1 year	unknown	No user rating, does not list who started the thread, doesn't list when the thread was started, blatant rip off digg with one tenth the features,

Reddit	reddit.com	Social News	when article was posted, who posted it, number of points the article has, number of comments the article has, where the article comes from, view all user comments, view all user posts	unknown	unknown	No user creation date, no user rating,
MetaFilter	metafilter.com	Social News	who submitted story, when it was submitted, how many comments it has, total user posts, total user comments, users friends and their contributions	7 years	unknown	no article ratings, no user ratings, site has so many different pieces of text hyperlink highlighted that its hard to distinguish reporting has potential issues with abuse, no total number of views so that the number of positive marks means something
ShoutWire	shoutwire.com	Social News	who posted it, when it was posted, number of comments, number of positive checks, ability to report articles, user join date	1.5 years	> 3000	no specific user rating, no real article rating, no total views so "pops" mean something
ClipMarks	clipmarks.com	Social News	who submitted, who submitted, number of votes, number of comments, article source, when member joined, total articles posted, total articles "popped" total number of "pops"	approx 1 year	unknown	no thread rating, no user rating, no user information, no user join date, no user information of any kind, doesn't list who posted the article
Tailrank	tailrank.com	Social News	when posted, links related to thread,	approx 1 year	unknown	
Now Public	nowpublic.com	Social News	who posted, when posted, number of comments, number of views, users recent stories, users favorite contributors	2 years	unknown	no user rating, no thread ratings,

Table 1-2 – Comparison of definitions in an attempt to redefine peer to peer network category names

Androutsellis-Theotokis, Spinellis, 2004	DEFINITIONS	COMMENTS	REFINED CONCEPTUALIZATION
Peer-to peer systems are distributed systems consisting of interconnected nodes, able to self-organize into network topologies with the purpose of sharing resources such as content, CPU cycles, storage, and bandwidth	A centralized(hybrid) peer-to-peer network in which user created video/audio/photo content is distributed to registered and non registered users alike through the use of a third party program or hosting website. Users are allowed personal accounts in which they can upload and share content, make comments, rate content, and establish an identity. The common goal of the video peer-to-peer network is to be the first to release new content thereby having the right to say "we had it first."	The definitions here are very similar, I think. Both the ACM and my definition talk about interconnected nodes (or users in my definition), to distribute content. The different I see is that their definition focuses on interconnected nodes to self organize. My network is interconnected nodes, but I do not think they self organize, since the organization would be done via the central server. I would disagree with them that all p2p networks are able to self organize.	A centralized peer-to-peer network consisting of nodes connecting to a centralized server with the purpose of content sharing of media files.
Peer-to peer systems are distributed systems consisting of interconnected nodes, able to self-organize into network topologies with the purpose of sharing resources such as content, CPU cycles, storage, and bandwidth	A decentralized(pure) peer-to-peer network in which users connect to other users directly through use of a third party program to share audio/video/software/games/ pornography, most of which is pirated. The audio peer-to-peer networks are plagued with information integrity problems, and contain no user or file rating systems, but do contain search engines. Audio networks are setup and taken down frequently due to copyright infringement problems, causing users to find newer and "more secure" options.	These definitions seem to match well. Both of them mention self organizing network topologies and distributed computing, with the purpose of sharing resources. The audio p2p networks are basically designed this way, and the ACM definition fits this specific category almost perfectly. The only point I would raise is about my information integrity. With the audio networks in particular, information integrity is a very large problem and definitely needs to be mentioned in the definition.	A decentralized peer-to-peer network consisting of interconnected nodes able to self organize with the purpose of trading audio, video, software, games and pornography, most of which is pirated, and where networks are plagued with information integrity problems due to the structure of the network.
Peer-to peer systems are distributed systems consisting of interconnected	Centralized (hybrid) peer-to-peer network in which users	I think our definitions are very different here, and I think that	A centralized peer-to-peer network consisting of a central

nodes, able to self-organize into network topologies with the purpose of sharing resources such as content, CPU cycles, storage, and bandwidth	work primarily with open source software distributions and distribute that software freely to anyone who wishes to use it. Consists of communities that, while they may be smaller in numbers, are generally older/well embedded, where information integrity problems are almost non existent.	a merger of the 2 would make a nice compromise and make a nice definition. I do not like how the ACM does not mention anything about information integrity, network sizes or ways of keeping users in line. As such, I feel that adding a sentence about information integrity will strengthen the definition.	server with many users contributing their content for the purpose of sharing their open source software, where information integrity is kept in check due to communities being older and well embedded.
Peer-to peer systems are distributed systems consisting of interconnected nodes, able to self-organize into network topologies with the purpose of sharing resources such as content, CPU cycles, storage, and bandwidth	Centralized (hybrid) peer-to-peer networks which rely on the use of third party websites and are driven largely by user created/submitted content. Integrity is kept in check mostly due to user profiles which allow for user tracking, and user voting which allows for popularity checks. Deals networks are not as plagued by the information integrity problem as the audio networks are, but still have more problems than the software networks do.	My definition really doesn't mention anything about what is shared, and it needs to. However, once again, the ACM definition mentions nothing on information integrity which I feel that it should. I also think that the ACM definition simply stating distributed systems is not specific enough, and only really applies to the audio networks.	Centralized peer-to-peer network that relies on nodes connecting to a centralized server for the purpose of sharing knowledge to other nodes, where information integrity is kept in check by a nodes peers.

Figure 1-1 – Conjoint Profile Analysis As Variables Are

Variable	Step 1	Step 1	Step 2	Step 2	Step 3	Step 3	Step 4	Step 4	Relative
	β	Z-Stat	β	Z-Stat	β	Z-Stat	β	Z-Stat	Importance
Constant	3.286	6.433	3.286	6.652	3.286	7.159	3.286	7.714	
Confdnce	.088	2.425	.088	2.507	.088	2.699	.088	2.908	
HoursDay	-.041	-1.092	-.041	-1.129	-.041	-1.215	-.041	-1.309	
UseFreq	.120	3.324	.120	3.437	.120	3.699	.120	3.986	
HowLong	-.016	-.445	-.016	-.461	-.016	-.496	-.016	-.534	
MIS	.118	1.660	.118	1.716	.118	1.847	.118	1.990	
NonMIS	.212	2.416	.212	2.498	.212	2.688	.212	2.897	
Dual	.171	3.985	.171	4.121	.171	4.435	.171	4.778	
Freshman	-.059	-1.389	-.059	-1.436	-.059	-1.545	-.059	-1.665	
Sophomore	-.165	-1.927	-.165	-1.993	-.165	-2.145	-.165	-2.311	
Junior	-.202	-2.447	-.202	-2.531	-.202	-2.724	-.202	-2.935	
Senior	-.221	-3.346	-.221	-3.460	-.221	-3.723	-.221	-4.012	
Age	.113	2.671	.113	2.762	.113	2.973	.113	3.203	
Gender	.033	1.003	.033	1.037	.033	1.116	.033	1.203	
CntrbRep			.218	7.532	.218	8.106	.218	8.733	3
RateCont			.127	4.377	.127	4.711	.127	5.076	6
FileQual					.328	12.193	.328	13.138	1
FileRate					.125	4.653	.125	5.013	7
ContIden							.138	5.532	4
ContUse							.085	3.405	8
FileUse							.129	5.162	5
P2PTrust							.246	9.844	2
EaseOfUse							.079	3.184	9
R2 adjusted	.034		.096		.220		.328		
F-Statistic (ΔR2)	3.884		8.659		18.868		24.920		
ΔR2 adjusted	.045		.064		.123		.110		

Figure 1-2 Survey Instruction Sheet

2007 Internet Peer-to-Peer (P2P) Survey

INSTRUCTIONS: Assume that you are searching Internet peer-to-peer (P2P) networks (e.g., Napster, You-Tube, Kazaa, Limewire, or Bittorrent) to find a copy of a $30 video game or movie DVD of interest to you. The tables in the survey describe hypothetical profiles of 12 different search results that you have obtained. Based on this information and your expertise, please answer the two questions next to each table below (by circling the shaded boxes).

Here, CONTRIBUTOR refers to the person who contributed the FILE that you are considering downloading. Assume that:

- You are completely anonymous
- The file is legal and takes 10 minutes to download (500 Megabytes in size)
- The P2P network has existed for 5 years.

The table below describes each characteristic in the search profile tables.

REFERENCE CARD		
	CONTRIBUTOR reputation	Rating of this contributor by other users of this P2P network.
	# of ratings for CONTRIBUTOR	How many previous downloaders have rated this contributor.
	FILE quality rating	How highly previous downloaders have rated this file.
	# of ratings for this FILE	How many previous downloaders have rated this file.
	P2P network trustworthiness	How much confidence you have that this P2P network is not flooded with fake/ malicious files (e.g., containing spyware or viruses).
	CONTRIBUTOR identifiability	The extent to which contributors' identity can be verified in this P2P network (e.g., by requiring registration, valid email addresses, etc.).
	File rating system usage	How extensively a file rating system is used in this P2P network (e.g., detailed comments left by previous downloaders).
	Contributor reputation system usage	How extensively a contributor reputation/ rating system is used in this P2P network.
	Ease of use of ratings	How easy it is to use the file and contributor rating system to make download decisions.

Figure 1-3 – Survey Instrument

2007 Internet Peer-to-Peer (P2P) Survey

Please answer the two questions next to each table below (by circling the shaded boxes).

Search Results # 1 of 12

CONTRIBUTOR reputation	**High**
# of ratings for CONTRIBUTOR	Low
CONTRIBUTOR identifiability	Low
Contributor reputation system usage	**High**
FILE quality rating	**High**
# of ratings for this FILE	**High**
File rating system usage	Low
P2P network trustworthiness	Low
Ease of use of ratings	Low

For downloading this file to my computer…
Risks Greatly Exceed Benefits 1 2 3 4 5 6 7 8 9 **Benefits** Greatly Exceed Risks
What is the likelihood that you will download this file?
Very **Low** 1 2 3 4 5 6 7 8 9 Very **High**

Search Results # 2 of 12

CONTRIBUTOR reputation	**High**
# of ratings for CONTRIBUTOR	Low
CONTRIBUTOR identifiability	Low
Contributor reputation system usage	Low
FILE quality rating	Low
# of ratings for this FILE	**High**
File rating system usage	**High**
P2P network trustworthiness	**High**
Ease of use of ratings	Low

For downloading this file to my computer…
Risks Greatly Exceed Benefits 1 2 3 4 5 6 7 8 9 **Benefits** Greatly Exceed Risks
What is the likelihood that you will download this file?
Very **Low** 1 2 3 4 5 6 7 8 9 Very **High**

Search Results # 3 of 12

CONTRIBUTOR reputation	Low
# of ratings for CONTRIBUTOR	**High**
CONTRIBUTOR identifiability	Low
Contributor reputation system usage	**High**
FILE quality rating	Low
# of ratings for this FILE	Low
File rating system usage	Low
P2P network trustworthiness	**High**
Ease of use of ratings	Low

For downloading this file to my computer…
Risks Greatly Exceed Benefits 1 2 3 4 5 6 7 8 9 **Benefits** Greatly Exceed Risks
What is the likelihood that you will download this file?
Very **Low** 1 2 3 4 5 6 7 8 9 Very **High**

Search Results # 4 of 12

CONTRIBUTOR reputation	Low
# of ratings for CONTRIBUTOR	**High**
CONTRIBUTOR identifiability	Low
Contributor reputation system usage	Low
FILE quality rating	**High**
# of ratings for this FILE	**High**
File rating system usage	Low
P2P network trustworthiness	Low
Ease of use of ratings	**High**

For downloading this file to my computer…
Risks Greatly Exceed Benefits 1 2 3 4 5 6 7 8 9 **Benefits** Greatly Exceed Risks
What is the likelihood that you will download this file?
Very **Low** 1 2 3 4 5 6 7 8 9 Very **High**

Search Results # 5 of 12

CONTRIBUTOR reputation	Low
# of ratings for CONTRIBUTOR	**High**
CONTRIBUTOR identifiability	**High**
Contributor reputation system usage	**High**
FILE quality rating	Low
# of ratings for this FILE	**High**
File rating system usage	**High**
P2P network trustworthiness	Low
Ease of use of ratings	Low

For downloading this file to my computer…
Risks Greatly Exceed Benefits 1 2 3 4 5 6 7 8 9 **Benefits** Greatly Exceed Risks
What is the likelihood that you will download this file?
Very **Low** 1 2 3 4 5 6 7 8 9 Very **High**

Search Results # 6 of 12

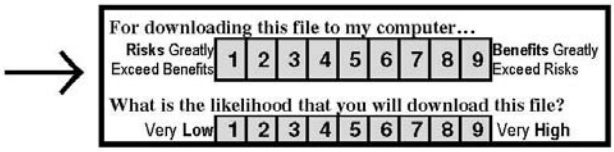
CONTRIBUTOR reputation	**High**
# of ratings for CONTRIBUTOR	Low
CONTRIBUTOR identifiability	**High**
Contributor reputation system usage	**High**
FILE quality rating	Low
# of ratings for this FILE	Low
File rating system usage	Low
P2P network trustworthiness	Low
Ease of use of ratings	**High**

For downloading this file to my computer…
Risks Greatly Exceed Benefits 1 2 3 4 5 6 7 8 9 **Benefits** Greatly Exceed Risks
What is the likelihood that you will download this file?
Very **Low** 1 2 3 4 5 6 7 8 9 Very **High**

CONTINUED ON REVERSE ►►►

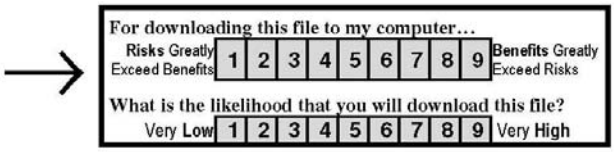
Search Results # 7 of 12

CONTRIBUTOR reputation	**High**
# of ratings for CONTRIBUTOR	**High**
CONTRIBUTOR identifiability	Low
Contributor reputation system usage	Low
FILE quality rating	Low
# of ratings for this FILE	Low
File rating system usage	**High**
P2P network trustworthiness	Low
Ease of use of ratings	**High**



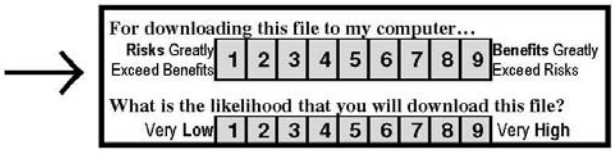
Search Results # 8 of 12

CONTRIBUTOR reputation	**High**
# of ratings for CONTRIBUTOR	**High**
CONTRIBUTOR identifiability	**High**
Contributor reputation system usage	Low
FILE quality rating	**High**
# of ratings for this FILE	Low
File rating system usage	Low
P2P network trustworthiness	**High**
Ease of use of ratings	Low



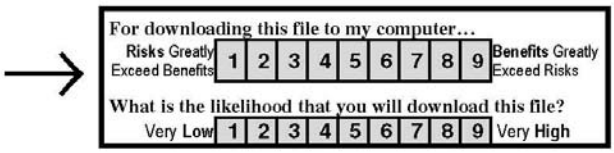
Search Results # 9 of 12

CONTRIBUTOR reputation	Low
# of ratings for CONTRIBUTOR	Low
CONTRIBUTOR identifiability	Low
Contributor reputation system usage	**High**
FILE quality rating	**High**
# of ratings for this FILE	Low
File rating system usage	**High**
P2P network trustworthiness	**High**
Ease of use of ratings	**High**



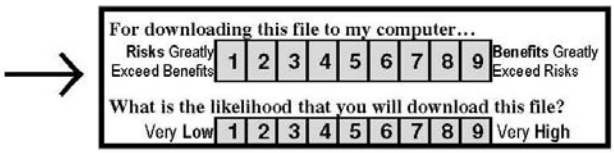
Search Results # 10 of 12

CONTRIBUTOR reputation	**High**
# of ratings for CONTRIBUTOR	**High**
CONTRIBUTOR identifiability	**High**
Contributor reputation system usage	**High**
FILE quality rating	**High**
# of ratings for this FILE	**High**
File rating system usage	**High**
P2P network trustworthiness	**High**
Ease of use of ratings	**High**



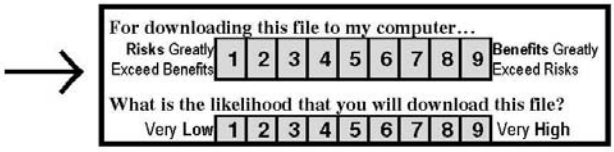
Search Results # 11 of 12

CONTRIBUTOR reputation	Low
# of ratings for CONTRIBUTOR	Low
CONTRIBUTOR identifiability	**High**
Contributor reputation system usage	Low
FILE quality rating	Low
# of ratings for this FILE	**High**
File rating system usage	Low
P2P network trustworthiness	**High**
Ease of use of ratings	**High**



Search Results # 12 of 12

CONTRIBUTOR reputation	Low
# of ratings for CONTRIBUTOR	Low
CONTRIBUTOR identifiability	**High**
Contributor reputation system usage	Low
FILE quality rating	**High**
# of ratings for this FILE	Low
File rating system usage	**High**
P2P network trustworthiness	Low
Ease of use of ratings	Low



How confident are you about your evaluation of the above tables (circle one)?

Little confidence	1	2	3	4	5	6	7	8	9	10	11	High confidence

Approximately, how many hours/day do you use the Internet (☑ one)? ☐0-1 ☐2-4 ☐5-7 ☐8+
How often do you use peer-to-peer networks? ☐Almost never ☐Occasionally ☐Regularly
How long have you used peer-to-peer networks? ☐Under a year ☐1-2 years ☐3-4 years ☐4+ years
Major (☑ one)? ☐MIS ☐Non-MIS ☐Dual major (with MIS) ➔**Status**? ☐Freshman ☐Sophomore ☐Junior ☐Senior
Age (years)? ☐18 or younger ☐19 ☐20 ☐21 ☐22 ☐23+
Gender (☑ one)? ☐Female ☐Male

Thank You ❖ **PLEASE RETAIN THE RED TICKET STUB** ❖

References

Aberer, Karl, and Zoran Despotovic. Managing Trust in a Peer-2-Peer Information System. Swiss Federal Institute of Technology. Lausanne, 2001. 1-8.

Androutsellis-Theotokis, Stephanos, and Diomidis Spinellis. "A Survey of Peer-to-Peer Content Distribution Technologies." ACM Computing Surveys ns 36 (2004): 335-371.

Aringhieri, R, E Damiani, S De Capitani Di Vimercati, and P Samarati. Assessing Efficiency of Trust Management in Peer-to-Peer Systems. Milan University. 1-10.

Caglj, Mario, Srdjan Capkun, and Jean-Pierre Hubaux. "Key Agreement in Peer-to-Peer Wireless Networks." Proceedings of the IEEE ns 94 (2006): 467-478.

Capkun, Srdjan, Jean-Pierre Hubaux, and Levente Buttyan. "Mobility Helps Peer-to-Peer Security." IEEE Transactions on Mobile Computing ns 5 (2006): 43-51.

Chien, Erin. Malicious Threats of Peer-to-Peer Networking. Symantec. Symantec, 2003. 1-12.

Chien, Erin. Malicious Threats of Peer-to-Peer Networking. Symantec. Symantec, 2003. 1-12.

Cooper, Brian F., and Hector Garcia-Molina. "Ad Hoc, Self-Supervising Peer-to-Peer Search Networks." ACM Transactions on Information Systems ns 23 (2006): 169-200.

Damiani, Emesto, Sabrina De Capitani Di Vimercati, Stefano Paraboschi, and Pierangela Samarati. "Managing and Sharing Servents' Reputations in P2P Systems." IEEE Transactions on Knowledge and Data Engineering ns 15 (2003): 840-854.

Feldman, Michal, and John Chuang. "Overcoming Free-Riding Behavior in Peer-to-Peer Systems." ACM SIGecom Exchanges ns 5 (2005): 41-50.

Feldman, Michal, Kevin Lai, Ion Stoica, and John Chuang. Robust Incentive Techniques for Peer-to-Peer Networks. 17 May 2004, ACM. New York: ACM, 2004.

Goel, Sharad, Mark Robson, Milo Polte, and Emin G. Sirer. Herbivore: a Scalable and Efficient Protocol for Anonymous Communication. Cornell University. Ithaca: Cornell. 1-17.

Granville, Lisandro Z., Diego M. De Rosa, Andre Panisson, Cristina Melchoirs, Maria J. Bosuiroli Almeida, and Liane M. Rockenbach Tarouco. "Managing Computer Networks Using Peer-to-Peer Technologies." IEEE Communications Magazine 1 Oct. 2005: 62-68.

Gupta, Minaxi, Paul Judge, and Mostafa Ammar. A Reputation System for Peer-to-Peer Networks. NOSSDAV, 1 June 2003, ACM. Monterey-California: ACM, 2003.

Hughes, Daniel, James Walkerdine, Geoff Coulson, and Stephen Gibson. "Peer-to-Peer: is Deviant Behavior the Norm on P2P File-Sharing Networks?" IEEE Distributed Systems Online ns 7 (2006): 1-11.

Kalafut, Andrew, Abhinav Acharya, and Minaxi Gupta. "A Study of Malware in Peer-to-Peer Networks." ACM IMC ns (2006): 1-6.

Kalafut, Andrew, Abhinav Acharya, and Minaxi Gupta. "A Study of Malware in Peer-to-Peer Networks." ACM IMC ns (2006): 1-6.

Kalafut, Andrew, Abhinav Acharya, and Minaxi Gupta. "A Study of Malware in Peer-to-Peer Networks." ACM IMC ns (2006): 1-6.

Kamvar, Sepandar D., Mario T. Schlosser, and Hector Garcia-Molina. "The EigenTrust Algorithm for Reputation Management in P2P Networks." ACM ns (2003): 1-12.

Kwok, S H., K Y. Chan, and Cheung Y. M. "A Server-Mediated Peer-to-Peer System." ACM SIGecom Exchanges ns 5 (2005): 38-47.

Lee, So Young, O-Hoon Kwon, Jong Kim, and Sung Je Hong. A Reputation Management System in Structured Peer-to-peer Networks. Pohang University of Science and Technology. 1-6.

Lu, Yi, Weichao Wang, Bharat Bhargava, and Dongyan Xy. "Trust-Based Privacy Preservation for Peer-to-Peer Data Sharing." IEEE Transactions on Systems, Man, and Cybernetics ns 36 (2006): 498-502.

Lua, Eng K., Jon Crowcroft, Marcelo Pias, Revi Sharma, and Steven Lim. "A Survey and Comparison of Peer-to-Peer Overlay Network Schemes." IEEE Communications Surveys and Tutorials 2005: 72-93.

Parno, Bryan, and Mema Roussopoulos. "Defending a P2P Digital Preservation System." IEEE Transactions on Dependable and Secure Computer ns 1 (2004): 209-222.

Parameswaran, Manoj, Anjana Susarla, and Andrew B. Whinston. "P2P Networking: an Information-Sharing Alternative." IEEE Distributed Systems (2001): 31-38.

Pitsilis, Georgios, and Lindsay Marshall. "A Trust-Enabled P2P Recommender System." 1-6.

Resnick, Paul, Richard Zeckhauser, Eric Friedman, and Ko Kuwabara. "Reputation Systems." Communications of the ACM ns 43 (2000): 45-48.

Rubenstein, Dan, and Sambit Sahu. "Can Unstructured P2P Protocols Survive Flash Crowds." IEEE/ACM Transactions on Networking ns 13 (2005): 501-512.

Santi, Paolo. "Topology Control in Wireless Ad Hoc and Sensor Networks." ACM Computing Surveys ns 37 (2005): 164-194.

Seedorf, Jan. "Security Challenges for Peer-to-Peer SIP." IEEE Networks ns (2006): 38-45.

Shieh, Shiuhpyng W., and Dan S. Wallach. "Ad Hoc and P2P Security." IEEE Internet Computing ns (2005): 14-15.

Song, Shanshan, Kai Hwang, Runfang Zhou, and Yu-Kwong Kwok. "Trusted P2P Transactions with Fuzzy Reputation Aggregation." IEEE Internet Computing ns (2005): 24-34.

Tiwana, Amrit. "Affinity to Infinity in Peer-to-Peer Knowledge Platforms." Communications of the ACM ns 46 (2003): 76-80.

Tiwana, Amrit, and Mark Keil. "Functionality Risk in Information Systems Development: an Empirical Investigation." IEEE Transactions on Engineering Management ns 53 (2006): 412-425.

Tiwana, Amrit, Jijie Wang, Mark Keil, and Punit Ahluwalia. "The Bounded Rationality Bias in Managerial Valuation of Real Options: Theory and Evidence From IT Projects." Decisions Sciences ns 38 (2007): 157-181.

Tiwana, Amrit, Mark Keil, and Robert G. Fichman. "Information Systems Project Continuation in Escalation Situations: a Real Options Model." Decisions Sciences ns 37 (2006): 357-391.

Wing, Yao, and Julita Vassileva. Trust and Reputation Model in Peer-to-Peer Networks. University of Saskatchewan. Saskatoon, 2003. 1-8.

Zhang, Xinwen, Songqing Chen, and Ravi Sandhu. "Authenticity and Integrity in P2P Systems." IEEE Internet Computing (2005): 42-49.